UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS

TESIS DOCTORAL

Desarrollo de una nueva generación de técnicas de imagen PET
Development of a new generation of PET imaging techniques

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Pablo Galve Lahoz

Directores

José Manuel Udías Moinelo
Joaquín López Herraiz
David Izquierdo García

Madrid

Universidad Complutense de Madrid
Facultad de Ciencias Físicas

Departamento de Estructura de la Materia, Física Térmica y Electrónica

DESARROLLO DE UNA NUEVA GENERACIÓN DE TÉCNICAS DE IMAGEN PET
DEVELOPMENT OF A NEW GENERATION OF PET IMAGING TECHNIQUES

MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR

Pablo Galve Lahoz

Directores:
José Manuel Udías Moinelo
Joaquín López Herraiz
David Izquierdo García

Madrid, 2022

A mis padres y hermanos,
por el apoyo incondicional.

A mi abuela,
por tus albóndigas.

A mi madre,
por enseñarme a luchar hasta el final.

> **If I have seen further,**
> **it is by standing on the shoulders of giants.**
>
> *Isaac Newton*

> **La vida es como un helado en un caluroso día de verano.**
>
> *Yo mismo*

Agradecimientos

Pues ya estaría. Allá van. Tropecientas páginas de PET, para el que las quiera. Pero este ha sido un trabajo de varios años que no se ha escrito solo. Muchos han sido los ilustres que se tropezaron con mi torpe paseo por la senda de la ciencia hasta hoy. Muchos a los que debo mucho, tanto en lo profesional como en lo personal. Pero empecemos por el principio. Después de tanto caminar, no pasa nada porque procrastinemos un poquito más, ¿no?

Corría el año 2015, cuando yo, un joven iluso que venía de un intenso erasmus y con un diploma de graduado en física recién sacado del horno, llegué a oír sobre un máster interuniversitario en física nuclear realmente atractivo. En mis oídos, esto sonaba a más noches de erasmus y proyecto Manhattan, así que empecé a fantasear con una carrera científica colmada de explosiones, locas aventuras nocturnas, radiación y éxitos. Hasta que un día alguien vino y me dijo:

—He estado pensando en tu TFM. Creo que vamos a hacer PET.

Era la voz de José Manuel poniéndome los pies en la tierra. Con él tenía mi beca de colaboración, tras ella vino mi TFM, y después la jugosa oportunidad de seguir trabajando en el GFN como niño Harvard. Y bueno, aunque no he logrado ser nominado al Nobel tantas veces como me hubiera gustado, creo que José Manuel siempre ha sabido encaminar mis esfuerzos en una dirección productiva. Más tarde fue Joaquín quien entró en escena para darme apoyo y seguir cultivando mis intereses por la imagen médica. A ambos, por transmitirme vuestra pasión por la ciencia, por los intensos debates de pizarra, por todo lo que me habéis enseñado, y sobre todo, por toda la confianza que habéis depositado en mí. Muchas gracias.

Pero sigamos, que esto acaba de empezar. Así, fui dando mis primeros pasitos por el GFN. Empecé desorientado, y como todo pollo de ciudad pequeña perdido en el loco corral que es Madrid, lo primero que hice fue buscar compañeros entre becarios nucleares. Con ellos he compartido clases, códigos, congresos, risas, papers, y todas las penurias que conlleva escribir una tesis. Los que han sido mis compañeros durante todo el doctorado, Alejandro y Amaia, y que seguirán siendo amigos allá donde caigan en el futuro. Gracias por haber estado ahí.

Siguiendo los recuerdos en mi camino por el GFN, atisbo canciones tarareadas a capella y olores a comida recalentada en el 239, eternos cafés al calorcito del Sol en la plaza de las ciencias, y penúltimas rondas en el Cántaro que siempre acababan en el "Suéltate el pelo". Y es que, antes de seguir, quiero aclarar que todos y cada uno de los días que he pasado en el GFN he tenido la fortuna de estar rodeado de gente que, a día de hoy, pongo en un pedestal. Empezaré por Víctor y Jaime, orador e intérprete, Don Quijote y Sancho... Castilla y La Mancha (el orden de los factores no altera el resultado). Con vosotros he compartido todo este camino, y, mientras escribo estas palabras, se me cae una lagrimilla pensando en vuestra salida del grupo. Sois unos artistas. Sigamos, que viene el gallinero, donde por aquel entonces se cacareaba más que se hablaba. De allí salieron May, Paula y Vicky, veteranas y referentes en la supervivencia predoctoral. Las primeras personas a las que vi doctorarse y sufrir con ello, y quienes me enseñaron que esto no iba solo de hacer tesis. Javi aka Ovéjix, y Sílvia, compañeros del máster nucelar donde empezó todo, y amigos incondicionales todo este camino. Afirmo contundente y sin dudar que el GFN tuvo mucha suerte cuando os dejasteis adoptar. Y por más GFNitos, que no se diga: Víctor Jr, Oli, Fer, Andrea, Clara, Dani, Raúl, Murias y Miguel. Cómplices todos de hazañas y volteretas a tutiplén. A Paloma, por su infinita sabiduría burocrática al servicio del GFN, y a Luis Mario, por mostrar siempre un apoyo simpático desde la cara senior del grupo. Cristina, Laura, Angel y Armando, con quienes tantas y tantas veces hemos luchado por un huequito para una veintena de personas en la cafetería. A Carmen, por su alegría y su amabilidad cada mañana cuando pasaba a arreglar nuestros desastrosos despachos. La vieja guardia del GFN, hoy ya leyendas en el pasillo de honor del grupo: Jaco, Leti, Bruno y Esther. Hurgando y hurgando, en mis primeros días encuentro a Mariano y Vadym, que donde quiera que estén (vendimiando, dando clases, o lo que sea), guardan mi aprecio y mi respeto. Y que no me quiero olvidar de otros tantos que estuvieron poquito tiempo, pero intenso, como Pauline, Julián, Araceli, Aidana, Estefanía, Miguel (alto ;-)), Jen, Fran, y Pedro. Y por supuesto, esa fuente vital de energía que nos toma el testigo, Tania, Nerea, Marcos, Esther, Gabriela, Víctor Noubillas, Nataly, y, como no, Briz. Sin restar a nadie, quiero cerrar este párrafo dando especial gracias al 236 de Vicky-Amaia, Paula, y Jaime. Con vosotros el despacho, además de ser un puesto de trabajo, ha sido mi particular Central Perk. Un lugar de encuentro con amigos, donde entre horas de trabajo tan pronto cabía un debate serio sobre la fecha exacta de la erupción del Vesubio, como una discusión sobre la cría de alpacas manchegas. A todos, os quiero.

No cabe menos que acordarme de Elsa en estos agradecimientos, por el apoyo incondicional, incluso cuando ya no tenía por qué darlo. Y, por supuesto, gracias a toda la humilde provincia de Soria. Quién me iba a decir a mí, forastero llegado de tierra extraña, que un pequeño Traspiés terminaría rodeándome de tanta buena gente que de una u otra manera acabó conmoviéndome. De sincero corazón, muchas gracias a todos los sorianos y sorianas que por más o por menos compartieron conmigo tantos buenos momentos, y en particular a ti, Elsa. Porque ese trocito que me disteis de Soria, lo llevaré por siempre en la memoria. Soria, que linda eres...

Keep going... Boston! Y tengo que mandar un abrazo enorme a David, que me acogió como mentor, y como amigo en mis tres meses en Boston. Gracias por formar parte de esta tesis, tu apoyo ha sido siempre de gran ayuda. And thanks to Ciprian as well, to offer me the chance to collaborate with his fancy PET helmet. And of course, thanks to all the Martinos people. Hanne, Mike Levine, Hassan, the other two Mikes :), Manuel, Lucy, and the rest of the team. I only regret I wasn't able to meet you again during my PhD (stupid Covid...).

And keep going... My Mountain Family!! Because talking about Boston means so many happy moments with Ed, Brian, Cozmo and Whisky. You gave me a home, when I was just looking for a house. I will always feel grateful to you (seriously, you are still invited to visit Teruel). Joining the family a month and a half later, Naiara! You were the perfect roomie, and a fundamental part of my short time around the US (I feel you already knew it). Te deseo lo mejor en tus próximos años como informática, y solo espero que en algún futuro no muy lejano podamos volver a encontrarnos.

Y qué sonrisa más tonta se me pone cuando me acuerdo de... ¡Los Fisiamigos! Con todas las fisifiestas, las fisiquedadas, los fisifindes, las fisicañas, los fisirreencuentros, las fisipedas, las fisiescapadas, las fisidespedidas de soltero, los fisibodorrios, los fisivermuts, la fisipatata cup, los fisirrelatos... Y seguro que me dejo un buen puñado de fisicosas. Quizás en esta vida llegue a arrepentirme de haber escogido física, pero nunca de haber tenido la suerte de conoceros y compartir con vosotros tantos buenos momentos.

Poniendo ahora mi vida en perspectiva, estoy seguro de que no hay nada bueno en una pandemia. Pero sí que quiero agradecer a toda la gente de Teruel que, tras haber vivido tantos años fuera, me acogió cuando tuve la oportunidad de volver a casa durante la misma. Allí nací y me crié, y sigo creciendo con los amigos de toda la vida cada vez que regreso. Gracias a Paquito y Carlos, por todos estos años de infinita complicidad. Ojalá os quede alguna visita a la capital en la recámara. A Miguel, Tintín, Luis, y todo el manisero, por haber crecido a mi vera. Y a todos los golfos y shatarreros, porque sois la mejor panda de liantes que podía pedir. Gracias a todos, por hacer que Teruel sea un sitio al que uno siempre quiere volver. Y ojalá, quizás, tal vez, en el que un día me pueda quedar.

¡Uf! Me toca hablar de mi casa, la de Madrid. De esta pequeña comuna que teníamos montada entre Jaime y Jorge, donde todo era de todos, salvo las tareas del hogar, que no eran de nadie (y el salón, que pertenecía al bosque salvaje de Jorge). En Antonia Ruiz Soro he tenido la mejor convivencia de mi vida. Gracias a ambos, y en especial a Jaime, que no solo me ha aguantado en casa, sino en el metro y en el despacho día tras día, y en tantas y tantas noches de cervezas, copas y lo que surja, hasta hacerme sentir hermanado suyo. A Jorge le deseo lo mejor en su decisión de asentar la cabeza con Heather (y que de ahí salga una invitación de boda). Y a Jaime, un próspero postdoc, mucha pizza, y mucha pasta. Estoy seguro de que los tres volveremos a coincidir en el tiempo y en el espacio.

Vamos llegando al final, y me toca hablar de la mujer más gata que he conocido. Chulería castiza, laísmo chulapo. Gracias Adriana, por aguantar mis desvaríos, e incluso haberme apoyado cuando yo estaba más en la Luna que en mi tesis. Cualquiera que sea el porvenir, gracias por compartir tu Madrid conmigo (y sí, reconozco que me gusta).

Ya casi cerrando, una vez dijo uno muy espabilado no se qué de que había podido ver muy lejos porque se había apoyado a hombros de gigantes. Pues bueno, que yo no seré tan espabilado ni espero que esta tesis vaya a tener gran relevancia, pero estoy seguro de que no hubiera completado este manuscrito sin el apoyo de todas las tesis del GFN que lo PETaron anteriormente, y que he tenido siempre a mano durante la escritura. A Samuel, Esther, Jacobo y Alex, gracias por vuestro trabajo.

Y por encima de todo, gracias a mi familia. A todos, primos, tíos, abuelos, hermanos y padres. A mi abuela, que me sigue mimando como si estuviera más cerca de la decena que de la treintena. A mis hermanos, por haber sido siempre una parte de mi vida. Y a mis padres, no solo por educarme en los valores de la constancia y el esfuerzo que me han llevado a donde estoy ahora, sino por ser el apoyo que siempre me sujeta cuando parece que todo se va a derrumbar.

Ya acabo, lo prometo. A los nucelares, a los compañeros de rugby, a los corredores, a Amacon (la pequeña tienda de ultramarinos de la esquina), etc. Cuatro páginas de agradecimientos, y seguro que me he dejado nombres... Os pido disculpas si no os habéis encontrado por aquí, pero sois tantos los que de una u otra manera habéis formado parte de estos años que me ha sido imposible hacer un repaso completo. Estéis o no explícitamente en estos agradecimientos,

¡¡¡GRACIAS!!!

Contents

Motivation and objectives of this thesis

Positron Emission Tomography (PET) [Cherry et al., 2012] is currently a fundamental diagnostic tool in medical fields such as Oncology, Neurology and Cardiology, besides it is widely extended in clinical and preclinical research studies. Looking back, since the first commercial scanner was released in 1978 [Phelps et al., 1977], many milestones in hardware development and image reconstruction methods have been achieved to put the technique on its current state. Historically, the size of PET scanners has increased steadily: from 1975 to 2000, the number of detector elements doubled every 24 months [Nutt, 2002]. Later the scanners were upgraded from 2D to 3D reconstruction [Michel Defrise and Paul Kinahan, 1998]. In the late 1990s the first multimodal PET/CT scanners combining anatomical information to supplement PET imaging appeared, followed by PET/MRI scanners in 2010 [Cal-Gonzalez et al., 2018a]. Time-of-flight has been another important addition to improve signal-to-noise performance [Lecoq et al., 2020, Surti and Karp, 2016]. Very recently, total-body clinical scanners, up to 2 meters long, have been developed [Nadig et al., 2021, Slart et al., 2021], and nonstandard geometries have been introduced [Catana, 2019b, Samanta et al., 2021, Tao et al., 2018, Yoshida et al., 2020]. Preclinical scanners, with submillimeter resolution required to resolve the organs of small animals, have also joined the trend [Grkovski et al., 2015, Perez-Benito et al., 2018, Qi et al., 2011, Udias et al., 2018]. Other advances have relied in new detectors [Slomka et al., 2016]. Many manufacturers incorporated multi-layered detectors with a narrow crystal pitch [Cheng et al., 2021, Kang et al., 2021, Mohammadi et al., 2017], or monolithical detectors to increase scanner resolution [Freire et al., 2021, Stockhoff et al., 2021].

The reconstruction methods have progressed along with the hardware, from simple analytic methods, such as filtered backprojection (FBP), to more accurate iterative algorithms [Iriarte et al., 2016, Qi and Leahy, 2006, Tong et al., 2010]. The maximum likelihood expectation maximization (MLEM) [Shepp and Vardi, 1982], a mainstream in PET reconstruction, has been constantly under scrutiny [Ahn and Fessler, 2003, Ahn et al., 2015, Alenius and Ruotsalainen, 1997, Browne and de Pierro, 1996, Green, 1990, Hudson and Larkin, 1994, Nuyts et al., 2002, Reader et al., 2011]. The recent revolution of deep learning algorithms based on convolutional neural networks have been incorporated to PET image processing and reconstruction as well [Arabi et al., 2021, Gong et al., 2021, Reader et al., 2021, Reader and Schramm, 2021]. Their current limitation are the large number of datasets required to train the networks, which are not always available.

Overall, the complexity of new generation PET scanners is continuously increasing. The larger amount of PET data collected requires more computational power to keep a reasonable duration of the clinical workflow. The maturity of GPU parallel computing in recent years made it possible speed-ups of several orders of magnitude with respect to multi-core, multi-CPU or even cluster computing. Monte Carlo simulations are a cornerstone in scanner design, but the intense computational needs and the difficulty of dealing with non-conventional geometries limited the performance of traditional simulation packages. It is also important to highlight that physical limits of PET must be addressed in full to profit from hardware advances, and novel reconstruction methods should be investigated.

In this context of PET effervescence, this thesis has been developed in the Nuclear Physics Group (GFN) of the Complutense University of Madrid, which, since the mid 2000s, is specializing in medical applications of nuclear physics. Specifically, the group experience in PET spans scanner design, development and performance evaluation [Vicente et al., 2013, 2007], image reconstruction [Cal-González et al., 2015b, España et al., 2007, Herraiz et al., 2011, 2006a, 2008, Lage et al., 2015, López-Montes et al., 2020], and PET MC simulations [Cal-González et al., 2013, 2015a, España et al., 2009, Lopez-Montes et al., 2019]. Previous theses developed in the group dealt with these topics [Cal-Gonzalez, 2014, España, 2009, Lopez-Montes, 2021, van Gómez López, 2021, Vicente, 2012]. In the framework of a close, long-standing and very fruitful collaboration with the Sedecal Medical Imaging (SMI) company, the family of SuperArgus PET scanners was released in 2015 [Udias et al., 2018]. These are a family of preclinical PET scanners with submilimeter resolution, unprecedented real time 3D-imaging, second to none in high rate capability and modular design to adapt to customer requirements, from very small to extended FOV, and standard as well as weird geometries. On the other hand, part of the work from this thesis was developed in collaboration with the Athinoula A. Martinos Center for Biomedical Imaging (Massachusetts General Hospital) and Harvard Medical School, where the author of this thesis completed a three-month internship in 2019. At Martinos center, the author was in close contact with current issues derived from the scatter correction in clinical scanners [Heußer et al., 2017, Lindemann et al., 2019] and the development of a state-of-the-art brain dedicated PET scanner with spherical geometry, the High Spatio-Temporal Resolution BrainPET (HSTR-BrainPET) [Catana, 2019a], which poses new challenges due to the size of the scanner (number of detectors and size of pixel elements), its geometry, and the simulations and reconstructions required for optimizing its design.

The main goal of this thesis was the development of a new generation of PET tools, covering the whole spectra of present needs of state-of-the-art PET scanners, from simulation to data correction and image reconstruction, exploiting the capabilities of current GPUs. We summarize them in the following list:

- The development of a fast and flexible reconstruction software based on the MAP-OSEM algorithm [Alenius and Ruotsalainen, 1997, Hudson and Larkin, 1994, Nuyts et al., 2002]. This tool must use a powerful framework to define arbitrary scanner geometries, and it must offer a modular structure which can be easily adapted to incorporate additional reconstruction algorithms.

- The study of novel methods to improve the current state-of-the-art in PET image quality, and its implementation in the reconstruction tools developed in the GFN.

- The development of a fast and flexible MC simulator. This simulator must satisfy the current needs of modeling modern scanners for multiple applications and optimizations, such as scatter correction or system response matrix (SRM) optimization.

Summary

Introduction and motivation

Positron Emission Tomography (PET) [Cherry et al., 2012] is a branch of nuclear medicine in which β^+ radiation is used to provide valuable information on physiological processes in the body. Since the first commercially available scanner was developed in 1978 [Phelps et al., 1977], important advances have been made in the technique to improve the image quality. As a result, the amount and complexity of data generated by new generations of PET scanners have drastically increased, requiring more computational power to keep the clinical workflow at reasonable duration, as well as improved models to make use of the increased resolution and improving corrections.

In this thesis, we have tackled the current needs with the development of different tools for PET image reconstruction and simulation on the basis of speed, flexibility, and accuracy. All the software and methods presented in this work have been accelerated with parallel computing on the graphical processing units (GPU), and they have been designed to adapt to any scanner design. In the following, we present a summary of every chapter of this thesis.

GPU-based PET image reconstruction software

In this chapter, we described the structure of the GPU-based PET image reconstruction software (GPU-PET reconstructor). We pursued an optimal trade-off between accuracy and flexibility, prioritizing speed performance whenever possible. The maximum *a posteriori* ordered subsets expectation maximization (MAP-OSEM) algorithm [Hudson and Larkin, 1994], with two different priors [Alenius and Ruotsalainen, 1997, Nuyts et al., 2002], was implemented. The system response matrix (SRM) was decomposed into easy GPU-parallelizable image filters and the forward/backward projection kernel. The main distinction of the GPU-PET is its input scanner and data format description, defined in two files with the crystal coordinates and the crystals-line of response (LOR) relation, gaining flexibility to work with any scanner geometry.

We show the results with simulated data from the Biograph mMR scanner [Delso et al., 2011] and a prototype spherical scanner, the High Spatio Temporal Resolution BrainPET (HSTR-BrainPET) [Catana, 2019a,b], and real data from the six-ring SuperArgus scanner (6R-SuperArgus) [Udias et al., 2018].

Improved Image Reconstruction with Super-iterations

A novel reconstruction algorithm to improve the image resolution beyond the current state-of-the-art was proposed [Galve et al., 2021c]. The algorithm was based on maximum-likelihood weights computed after the image reconstruction to achieve the optimal distribution of the counts measured in each LOR among several subLORs. We called this step a super-iteration. We implemented the algorithm with the MAP-OSEM in the GPU-PET reconstructor, but it was proposed in a general framework, compatible with any other reconstruction algorithm. We showed the results for real data in the four-ring SuperArgus scanner (4R-SuperArgus) [Udias et al., 2018]. A quantitative improvement of 10% in the recovery coefficients (RC) for the NEMA NU4-2008 image quality phantom [National Electrical Manufacturers Association, 2008] for equivalent noise level was achieved, and visual improvement in the peak-to-valley ratio of a cold Derenzo phantom and a mouse heart is shown.

Ultra-fast Monte Carlo PET simulator

In this chapter, we developed a new Monte Carlo (MC) PET simulator based on GPU computing to accelerate the performance of the whole simulation process, the Ultra-fast Monte Carlo PET simulator (UMC-PET simulator). The simulator was easily adaptable to any scanner geometry and a wide variety of detectors based on a voxelized description. The code accuracy was validated against PeneloPET [España et al., 2009, Lopez-Montes et al., 2019] working on a single central processing unit (CPU), achieving similar results 2000 times faster.

We also implemented the UMC-PET simulator for fully MC scatter correction with real data from the Biograph mMR scanner using the GPU-PET reconstructor. We compared the results against the single scatter simulation (SSS) [Ollinger, 1996, Watson, 2000, Watson et al., 1996] estimation obtained with the scanner software, and a hybrid approach combining the UMC-PET scatter scale with the SSS distribution. The UMC-PET outperformed the SSS, eliminating the halo artifacts around the bladder in high-contrast images.

Full integrated tool: UMC-PET improved reconstruction

We implemented a hybrid reconstruction for the MAP-OSEM algorithm [Hudson and Larkin, 1994], combining a forward projection step based on the UMC-PET simulator and a backward projection step based on the SRM described in the GPU-PET reconstructor. This approach included the physical model, detector response, scatter and attenuation based on the MC estimation. We also compared against a decomposition of the MC projection with symmetries to increase the statistics, and separate scatter and attenuation estimation. We showed simulation results of a Derenzo phantom in the two-ring Argus scanner (2R-Argus), and quantitative measurements of recovery coefficients (RC), spill over ratio (SOR), and noise in real data of the image quality phantom of the NEMA NU4-2008 [National Electrical Manufacturers Association, 2008] acquired in the 6R-SuperArgus scanner [Udias et al., 2018]. The best RC values were obtained for the MC projection with symmetries (40.2% for the 1 mm rod, against 33.5% for the UMC-PET reconstructor), whereas the best SOR values were achieved by the fully MC projection method (2.53% for the water cylinder, against 3.49% for the UMC-PET). The computational burden of the approaches proposed is too high for its practical use (12 hours for the simulation and 4 hours for the reconstruction), but it can be used as a gold standard reference for concrete cases.

Resumen en Castellano

Introducción y motivación

La Tomografía por Emisión de Positrones (PET) [Cherry et al., 2012] es una rama de la medicina nuclear en la que la radiación β^+ se utiliza para obtener valiosa información sobre procesos fisiológicos en el organismo. Desde que el primer escáner PET comercial fuera desarrollado en 1978 [Phelps et al., 1977], se han hecho importantes avances en la técnica para mejorar la calidad de imagen. Como resultado, la cantidad y complejidad de los datos generados por las nuevas generaciones de escáneres PET ha aumentado drásticamente, requiriendo de una mayor capacidad computacional para poder mantener los protocolos clínicos en tiempos razonables, así como de modelos mejorados para aprovechar el aumento en resolución y mejorar correcciones en los datos.

En esta tesis hemos abordado las necesidades actuales del PET con el desarrollo de herramientas de reconstrucción y simulación PET, fundamentadas en la búsqueda de velocidad, flexibilidad, y precisión. Todos los software y métodos presentados han sido acelerados por medio de GPUs, y han sido diseñados para adaptarse al diseño de cualquier escáner. A continuación, presentamos un resumen de cada capítulo de esta tesis.

Un software de reconstrucción de imagen basado en GPU

En este capítulo describimos la estructura de un software de reconstrucción de imagen basado en unidades de procesamiento gráfico (GPU; reconstructor GPU-PET). Hemos apostado por el equilibrio entre precisión y flexibilidad, priorizando la velocidad siempre que ha sido posible. El algoritmo de maximización de la expectativa con subconjuntos ordenados y máximos *a posteriori* (MAP-OSEM) [Hudson and Larkin, 1994], con dos funciones de prioridad [Alenius and Ruotsalainen, 1997, Nuyts et al., 2002], ha sido implementado. La matriz de respuesta del sistema (SRM) ha sido descompuesta en filtros de imagen altamente paralelizables en GPU y en la rutina de proyección/reproyeccción. La mayor distinción de GPU-PET es la descripción de escáner y del formato de los datos, definidos en dos ficheros con las coordenadas de los cristales y la relación entre cristales y líneas de respuesta (LOR), ganando la flexibilidad para trabajar con cualquier geometría.

Se muestran los resultados de datos simulados en el escáner Biograph mMR [Delso et al., 2011] y en un prototipo de escáner esférico, el BrainPET de Alta Resolución Espacio Temporal (HSTR-BrainPET) [Catana, 2019a,b], y datos reales del escáner SuperArgus de seis anillos (6R-SuperArgus) [Udias et al., 2018].

Mejora de la reconstrucción de imágenes por Super-iteraciones

Se ha propuesto un novedoso algoritmo de reconstrucción para mejorar el estado del arte de la resolución PET [Galve et al., 2021c]. El algoritmo se basa en pesos de máxima verosimilitud calculados tras la reconstrucción de la imagen para obtener una distribución óptima del número de cuentas medidas en cada LOR entre diferentes subLORs. Llamamos super-iteración a este paso. Hemos implementado el algoritmo con el MAP-OSEM en el reconstructor GPU-PET, pero se ha descrito en un marco general compatible con cualquier otro método de reconstrucción. Se muestran resultados de datos reales en el escáner SuperArgus de cuatro anillos (4R-SuperArgus) [Udias et al., 2018]. Se ha obtenido una mejora cuantitativa del 10% en los factores de recuperación (RC) para el fantoma de calidad de imagen NEMA NU4-2008 [National Electrical Manufacturers Association, 2008] para un nivel de ruido equivalente, y se puede observar una mejora en el ratio pico-valle de un Derenzo frío y en el corazón de un ratón.

Simulador Monte Carlo PET ultra-rápido

En este capítulo, hemos diseñado un nuevo simulador PET Monte Carlo (MC) basado en programación en GPU para acelerar el rendimiento de todo el proceso de simulación, el simulador Monte Carlo PET ultra-rápido (simulador UMC-PET). Este simulador se adapta fácilmente a cualquier geometría y detector basados en una descripción voxelizada. El código se ha validado contra PeneloPET [España et al., 2009, Lopez-Montes et al., 2019] trabajando en una única unidad central de procesos (CPU), obteniendo resultados similares más de 2000 veces más rápido.

También hemos implementado el simulador UMC-PET para hacer corrección de dispersión puramente MC con datos del escáner Biograph mMR. Hemos comparado los resultados con la estimación del software del escáner por medio de simulación de dispersión única (SSS) [Ollinger, 1996, Watson, 2000, Watson et al., 1996], y un método híbrido combinando el escalado del UMC-PET con la distribución SSS. UMC-PET superó al SSS, eliminando artefactos de halo en la imagen alrededor de la vejiga en imágenes con alto contraste.

Herramienta integrada: reconstrucción mejorada con UMC-PET

Hemos implementado una reconstrucción híbrida para el algoritmo MAP-OSEM [Hudson and Larkin, 1994], combinando la proyección basada en el simulador UMC-PET y la reproyección basada en la SRM descrita en el reconstructor GPU-PET. Este enfoque incluye el modelado físico, la respuesta de los detectores, la dispersión y la atenuación basados en los cálculos MC. También comparamos contra una descomposición de la proyección MC con simetrías para incrementar la estadística, por separado de la contribución de la dispersión y la atenuación. Se muestran resultados de simulaciones de un fantoma Derenzo en el escáner Argus de dos anillos (2R-Argus), y medidas cuantitativas de los coeficientes de recuperación (RC), ratio de *spill over* (SOR), y el ruido con datos reales de un fantoma de calidad de imagen del NEMA NU4-2008 [National Electrical Manufacturers Association, 2008] adquirido en el escáner 6R-SuperArgus [Udias et al., 2018]. Los mejores valores de RC se obtuvieron para las proyecciones MC con simetrías (llegando a un 40.2% en el cilindro de 1 mm, frente a un 33.5% para el reconstructor GPU-PET), mientras que el mejor valor para el SOR se ha logrado con las proyecciones puras MC (2.53% en el cilindro de agua, frente a un 3.49% para el reconstructor GPU-PET). El coste computacional del método propuesto es demasiado alto para su uso diario (12 horas para la simulación y 4 horas para la reconstrucción), pero este puede llevarse a la práctica como método de referencia en casos concretos.

Chapter 1

Introduction

In the 1970s, the progress of computers revolutionized medical imaging with the implementation of tomography [Bradley, 2008], e.g. computerized tomogaphy (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single photon computerized tomography (SPECT). Among these techniques, PET and SPECT belong to the branch of molecular imaging techniques, in which a compound labeled with a radionuclide is injected into the patient body. We call this radiolabeled compound the radiopharmaceutical or radiotracer (or simply tracer). With the aid of an external gamma camera, we can detect the x-rays and gamma photons emitted that exit the patient body. When a sufficient number of emission events are accumulated, we can reconstruct the map of the tracer distribution, providing valuable biological information. Molecular imaging has become an essential tool to identify tumors in medicine [Papathanassiou et al., 2009], besides its application in other fields such as Neurology or Cardiology [Davis et al., 2020, Nutt, 2002]. Early this century, these methodologies have been combined with CT, and later on with MRI scanners, improving the diagnosis thanks to the addition of anatomical information [Cal-Gonzalez et al., 2018a].

In the case of PET [Cherry et al., 2012], which is the object of study in this thesis, the employed radionuclides have the particular characteristic of emitting positrons (they may also emit x-rays and gamma photons). The positrons commonly annihilate with a free electron in the medium after losing most of their energy describing a path that is from <1 mm up to few cm apart from the decay point, usually known as positron range. The annihilation results most often in two antiparallel gamma rays of 511 keV. The PET scanner is conformed by position sensitive detectors covering a wide region of

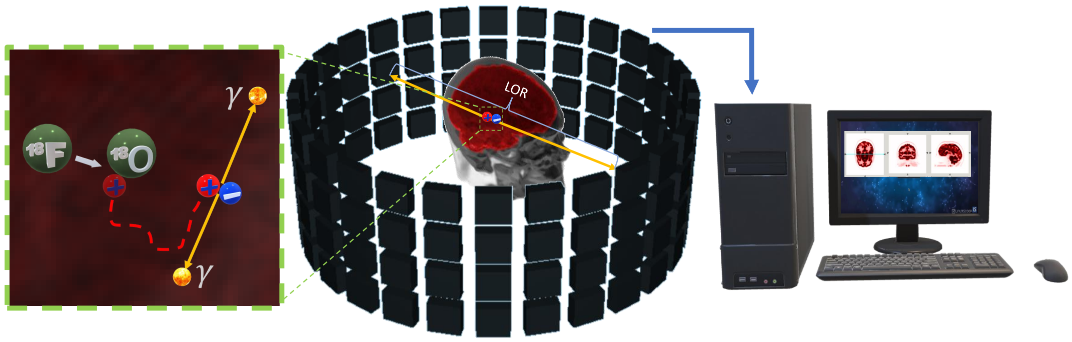


Figure 1.1: Scheme of the PET image formation. The radionuclide β^+ decay (in this case $^{18}F \rightarrow ^{18}O$), positron range, and annihilation are represented in the left hand side. The photons are detected in coincidence and processed by the scanner (central panel), and finally the data are reconstructed in a computer.

the patient. Whenever a photon successfully reaches the PET detectors, a single output signal is generated with a given time stamp. The scanner will search for temporal correlations among the individual detections (single events or simply *singles*) within a time window to identify pairs of annihilation photons, so called coincidence events. We call line of response (LOR) the physical line defined by the two single detection points in the scanner, which in principle contains the position of the annihilation event. A computer will process the data generated, getting advantage of the spatial and temporal correlations of the detected events to reconstruct the activity map. In figure 1.1 we show a scheme of the PET image formation process.

In this chapter we review the basics of PET and the main topics related to the work developed in the thesis. We start reviewing the physics behind PET in section 1.1, PET detectors in section 1.2, and PET acquisitions in section 1.3. In section 1.4 we talk about the main reconstruction algorithms applied in PET imaging and typical PET corrections. We describe the basics of Monte Carlo methods in section 1.5. We close the chapter with the keys of GPU computing and its applications in section 1.6.

1.1 The physics behind PET

We briefly summarized the main physical processes in the introduction of this chapter. In this section we will explain in more detail the β^+ decay and the physics associated to the positron emission and annihilation, and the main interactions suffered by gamma photons.

1.1.1 β^+ nuclear decay and positron emission

In the case of β-decay, either a proton or a neutron transforms into each other respectively, emitting a positron (β^+-decay) or an electron (β^--decay), and an associated neutrino or antineutrino [Krane, 1988],

$$\begin{aligned} p^+ &\rightarrow n + e^+ + \nu_e \\ n &\rightarrow p^+ + e^- + \bar{\nu}_e, \end{aligned} \tag{1.1}$$

where p^+ is the proton, n is the neutron, e^+ is the positron, e^- is the electron, and ν_e is the electronic neutrino (and antineutrino $\bar{\nu}_e$). These are the typical decay channels in proton-rich and neutron-rich nuclei respectively, besides, the β^+ decay cannot happen to free protons. From now on, we will focus on the β^+ decay,

$$^A_ZX \rightarrow ^A_{Z-1} X + e^+ + \nu_e, \tag{1.2}$$

where A is the mass number representing the total number of neutrons and protons, and Z is the atomic number representing the number of protons. ^A_ZX is the parent nucleus, and $^A_{Z-1}X$ the daughter nucleus. It is important to mention the electron capture (EC) decay channel, in which an electron from the inner shells of the atom is absorbed changing a proton into a neutron and emitting a neutrino,

$$^A_ZX + e^- \rightarrow ^A_{Z-1} X + \nu_e. \tag{1.3}$$

All the nuclei subject to decay with the β^+ channel may also decay via EC. We call the probability of events for each decay channel the Branching ratio.

The energy released in any of the β-decay channels depends on the decaying nucleus and it is shared between the emitted particles, and thus the positron has an energy that follows a characteristic energy spectrum. In figure 1.2 we can see an example of the β-spectrum for a ^{64}Cu nucleus. In the case of the electron spectrum (left hand side), there is a Coulomb energy barrier because of the opposite charges of the nucleus and the electron, and therefore we find a higher probability of emitting electrons with low energies.

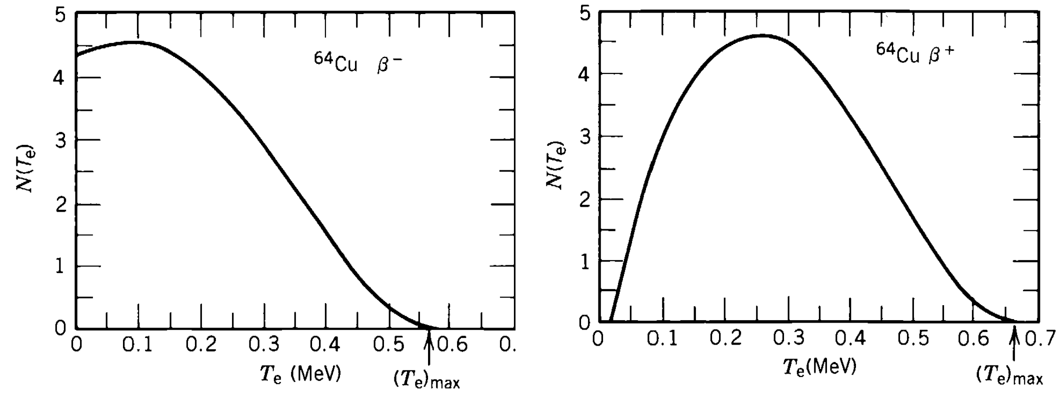


Figure 1.2: Kinetic energy spectrum ($N(T_e)$) of the released electron (left) and positron (right), in the β- and β+-decays of the ^{64}Cu [Krane, 1988]. The maximum energy released is marked as $(T_e)_{max}$.

In the following subsections, we briefly summarize two of the main limiting factors of the PET resolution derived from the nature of the β-decay: the positron range and the photon non-collinearity of annihilation photons.

1.1.1.1 Positron Range

Emitted positrons, being antimatter, will annihilate with an electron of the surrounding medium. As the positron annihilation mostly occurs at thermal energies (a few eV, much less than their rest energy [Berko and Hereford, 1956]), and the positron is released with a significant kinetic energy (right hand side of figure 1.2), this particle travels through the medium until it loses most of its kinetic energy via collisions with electrons and Bremsstrahlung radiation [Knoll, 2011]. We call the distance from the decay point to the annihilation point the positron range, and it is one of the fundamental resolution limits in PET [Alva-Sánchez et al., 2016, Cal-González et al., 2013, Cal-Gonzalez et al., 2009, Park et al., 2007, Peng and Levin, 2012]. In PET we simply refer to the positron range as the positron range average blurring caused by this factor in the decay image. This effect varies with the energy spectrum of emission and the interactions with the medium, i.e. the decaying radionuclide and the density composition of the surrounding materials. The kinetic energy distribution of the positron (figure 1.2) and the possible non-homogeneous material distribution is the cause that the range distribution is not Gaussian.

1.1.1.2 Non-collinearity

The second fundamental detriment in image quality caused by the β-decay physics is the non-collinearity of the annihilation photons. Because of the kinetic energy of the positron and the electrons in the medium at the moment of annihilation, the momentum conservation law requires that the relative angle of the emitted photons is not exactly 180º, but it has a Gaussian full width at half maximum (FWHM=$\Delta\phi$) proportional to the positron momentum [Erdman, 1955],

$$\Delta\phi = \frac{p_{e^+}}{m_e c}. \tag{1.4}$$

In [DeBenedetti et al., 1950], the experimental FWHM of this distribution was measured to be around 0.4º and 0.5º. However, since part of the total momentum is provided by the electron, this deviation may depend on the medium [do Nascimento et al., 2013].

1.1.2 Radiation-matter interaction

The annihilation photons may suffer different interactions when they travel through the patient and the detectors. As a result, they may be absorbed via photoelectric effect or they may change direction and lose part of their total energy via Compton or Rayleigh scatter. In PET, if either one of the annihilation photons interacts with the patient they will not be detected within the energy window or they will be detected in the wrong LOR, phenomenons that we refer as attenuation and scatter respectively. For a given photon energy, the probability of suffering an interaction increases exponentially with the path length d and the linear attenuation coefficient (μ) of the given material. We can define the attenuation of a photon beam of intensity I as

$$I(d) = I_0 e^{-\mu d}, \tag{1.5}$$

where I_0 is the initial intensity. The ratio I/I_0 is what we usually call the attenuation factor. The importance of attenuation in PET is described in detail in section 1.4.3.3.

For the energies of few keV to few MeV, the most probable interactions are coherent scatter (Rayleigh scatter), incoherent scatter (Compton scatter), photoelectric absorption, and pair production (for energies higher than 1.022 MeV) [Knoll, 2011]. The total

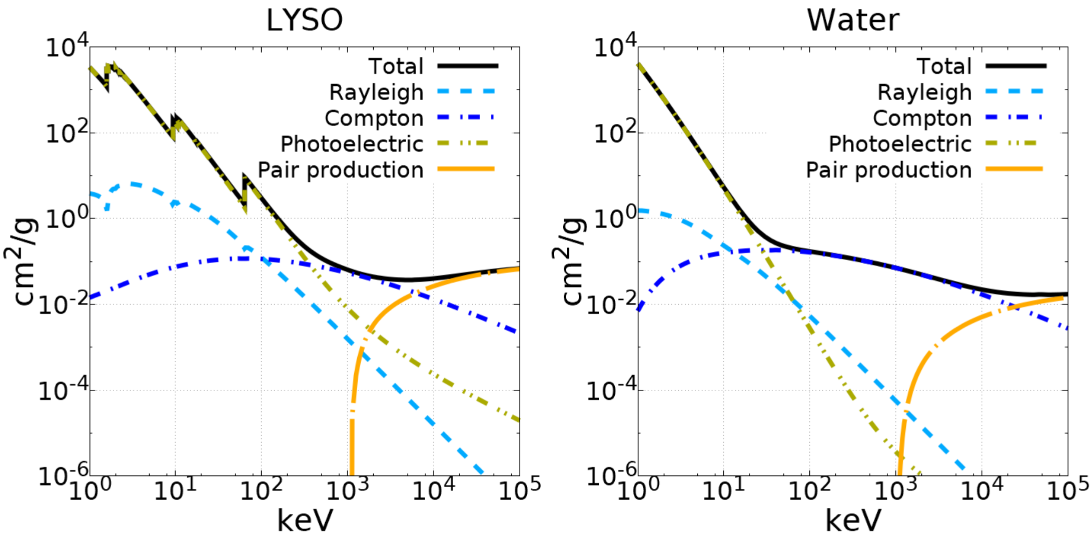


Figure 1.3: Photon cross sections for LYSO (left) and water (right) in the energy range of 1 keV to 100 MeV. We show the coefficients for Rayleigh scatter, Compton scatter, photoelectric absorption, pair production, and the total coefficients (calculated with the PENELOPE package [Agency, 2019]).

attenuation factor is the sum of each independent interaction attenuation factor and it depends on the photon energy,

$$\mu(E) = \mu_R(E) + \mu_C(E) + \mu_{PE}(E) + \mu_{PP}(E), \tag{1.6}$$

where μ_R, μ_C, μ_{PE} and μ_{PP} are respectively the attenuation factors of Rayleigh scatter, Compton scatter, photoelectric absorption and pair production.

In nuclear physics, we usually talk about the cross section (σ) instead of linear attenuation. The linear attenuation is equal to the cross section multiplied by the material density. In figure 1.3 we show the cross section of lutetium-yttrium orthosilicate (LYSO, a typical material for scintillation detectors in PET, see section 1.2) and water in the range of 1 keV up to 100 MeV. For 511 keV photons, 62.6% of the interactions in LYSO will undergo Compton scatter, and 32.0% will undergo photoelectric effect. In the case of water, that is much closer to the biological tissues, 99.8% of the interactions are Compton scatter.

Now we briefly explain the main characteristics of each interaction:

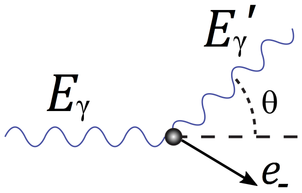


Figure 1.4: Schematic representation of the Compton scatter. E_γ represents the energy of the incident photon, E'_γ is the energy of the photon scattered an angle θ by the electron.

- **Rayleigh scatter**. This is the case of an elastic interaction, changing the photon direction without losing energy [Knoll, 2011]. This interaction is not dominant for the energies of interest in PET, but it might cause relevant dispersion at low energies (see figure 1.3).

- **Compton scatter**. In the Compton scatter a photon interacts with an electron from the medium, exchanging energy and momentum (see figure 1.4). In the laboratory center of reference, with the electron at rest, given a dispersion angle θ and initial energy E_γ for the photon we can estimate its energy after the collision (E'_γ) [Knoll, 2011],

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c^2}(1 - cos\theta)}, \tag{1.7}$$

where $m_e c^2$ is the electron rest energy. Maximum energy transfer occurs for $\theta = 180^o$.

The angular distribution of Compton scattering can be estimated from the Klein-Nishina formula. However, this is a rough approximation since atomic electrons are not at rest in reality, and electrons are bounded to the atomic shells [Agency, 2019]. The cross section of Compton scatter σ_C is energy dependant and proportional to the atomic number Z,

$$\sigma_C(E_\gamma, Z) = f(E_\gamma) \cdot Z. \tag{1.8}$$

Compton interaction is one of the main contributions to photon interactions at 511 keV (figure 1.3), and scattered photons may be detected in wrong LORs causing image deterioration. In section 1.4.3.4 we will discuss in deeper detail about this effect.

- **Photoelectric effect**. We call photoelectric absorption when an electron absorbs the total energy of the photon [Knoll, 2011]. The so called photoelectron is released from the atomic shell with an energy equal to the difference between the photon energy (E_γ) and the electron binding energy (E_B),

  $$E_{e^-} = E_\gamma - E_B. \tag{1.9}$$

  The most probable electrons to absorb a photon are in the outer shells, usually in the K shell. The binding energy is up to a few keV, therefore gamma photons will pass most of their energy to the photoelectron. Moreover, x-rays may be emitted in the electronic transitions from inner shells or Auger electrons.

  There is not a general expression for the photoelectric cross section, but it might be simplified in the following relation [Knoll, 2011]:

  $$\sigma_{PE} \propto \frac{Z^n}{E_\gamma^{3.5}}, \tag{1.10}$$

  where Z is the atomic number, and n is a number between 4 and 5.

  The photoelectric effect is, together with the Compton scatter, the most important contribution to the gamma photons interaction of 511 keV in the detector materials (in the case of LYSO, it represents 32.0% of the interactions; see figure 1.3). This is the most desired interaction in the PET detectors, since the photons are fully absorbed in a single interaction. Manufacturers search for materials with high Z number in our detectors to increase the probability of photoelectric absorption.

- **Pair production**. One last interaction is the pair production. This process is inverse to the photon annihilation. It consists in the transformation of a photon to give raise to an electron-positron couple. The process requires high energy photons of more than twice the electron rest mass (1022 keV), and in consequence it is not of high relevance in PET (the annihilation photons have 511 keV).

In figure 1.5 we show the relative relevance of each interaction for different energies and atomic numbers.

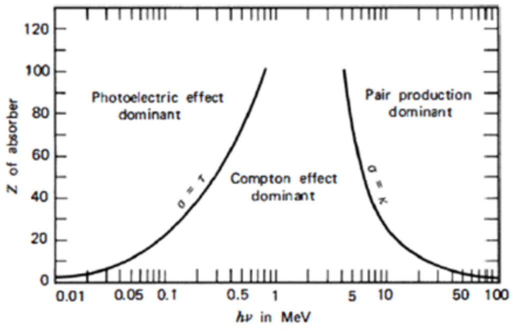


Figure 1.5: Comparison of the relative relevance of each interaction component with the energy dependence of gamma photons (from few keV up to 100 MeV) and the atomic number of the medium. The solid lines represent equality between the components. Image taken from [Knoll, 2011].

1.1.3 PET radionuclides

As it was aforementioned before, in PET we need a β^+ radionuclide that ideally emits a single positron to perform imaging (usually, additional gamma rays are emitted) [Bailey et al., 2005, Cherry et al., 2012]. Short half-lives are preferred (up to a few hours) for safety reasons and higher comfort for the patients. However, when the half-lives are of the order of a few minutes, the radionuclides can not be transported (such as the ^{11}C, the ^{13}N, or the ^{15}O), and they must be generated on-site in cyclotrons or in generators (chemically extracted from a long-life parent nucleus). Furthermore, the radionuclide must accomplish the appropriate chemical qualities, either being participant of metabolic processes or being suitable for binding with active biological molecules.

In table 1.1, we show the main properties of commonly used PET tracers. Among them, ^{18}F, ^{11}C, ^{13}N, and ^{15}O are the most common ones since they better match the aforementioned conditions.

Oncology is the most benefited clinical area from PET. The most common radiopharmaceutical is ^{18}F-FDG because tumor cells have an accelerated metabolism that requires a higher uptake of glucose. ^{18}F also has an adequate half-life to be transported from the production cyclotron center to surrounding hospitals.

In this thesis, all the real acquisitions were performed with ^{18}F-FDG.

Radionuclide	Half-life	Production	β^+ Branching ratio
^{18}F	109.8 min	Cyclotron	96.7
^{11}C	20.5 min	Cyclotron	99.8
^{13}N	10.0 min	Cyclotron	99.8
^{15}O	2.0 min	Cyclotron	99.9
^{68}Ga	67.7 min	Generator (^{68}Ge)	89.1
^{82}Rb	1.3 min	Generator (^{82}Sr)	95.4
^{124}I	4.2 days	Cyclotron	22.7
^{64}Cu	12.8 hours	Cyclotron	17.6
^{76}Br	16.2 hours	Cyclotron	55.0
^{94m}Tc	52.0 min	Cyclotron	70.2
^{86}Y	14.7 hours	Cyclotron	31.9

Table 1.1: Half-life, production method, and β^+ branching ratio of common PET radionuclides [Bailey et al., 2005, Cherry et al., 2012].

1.2 PET detectors

The typical detectors utilized in PET have been adopted from traditional nuclear physics to optimize gamma photons detection with high sensitivity, fast response and reasonable energy resolution. We introduce here the concepts of efficiency, energy resolution, and time resolution [Knoll, 2011].

- **Efficiency**. The efficiency of a detector is defined as the ratio between the detected photons and the emitted ones. We can differentiate between the geometrical efficiency and the intrinsic efficiency. The **geometrical efficiency** is given by solid angle coverage of the detector and the source. In PET scanners, the detectors are assembled to cover the highest possible solid angle, usually in cylindrical geometry to facilitate the patient insertion [Delso et al., 2011, Grkovski et al., 2015, Jakoby et al., 2009, Kang et al., 2021, Nadig et al., 2021, Wang et al., 2006], although other morphologies have also been studied [Perez-Benito et al., 2018, Qi et al., 2011, Samanta et al., 2021, Yoshida et al., 2020]. The **intrinsic efficiency** is given by the attenuation coefficient of the detector material and its length, and the electronic response of the detector. In PET, a higher sensitivity will reduce the activity needed to obtain similar image quality.

- **Energy resolution**. As a concept, the energy resolution is the minimum energy difference that your detector can distinguish. It can be measured from the energy spectrum generated by a monoenergetic source. In general, the energy resolution is a function of the energy, and it is usually given as a percentage ratio for a reference energy value ($R(\%) = FWHM/E$).
- **Time resolution**. Equivalently, we define the time resolution as the minimum time interval between two contiguous events that our detector is able to distinguish. The detector resolution is a convolution of all the components that conform it. In PET, we are interested in the **coincidence time resolution**, which can be measured from time difference histogram of many coincidence events.

Good efficiency improves the statistics in our detectors, the energy resolution is needed to discard scatter events (see section 1.4.3.4), and fast time resolution is needed to avoid random coincidences (see section 1.4.3.5) and obtain time-of-flight information (TOF) [Lecoq et al., 2020, Surti and Karp, 2016, Vandenberghe et al., 2016]. The scintillation detectors match these properties with a reasonable trade-off between efficiency, energy resolution, and time resolution, hence they have been widely implemented in PET [Lewellen, 2008]. In this section, we will focus on their main components: the scintillation crystal, the photosensor, and the electronics circuit. The scintillation crystal is responsible of "stopping" the gamma ray and for generating a light response. The photosensor is optically coupled to the crystal to collect the light and convert it into an electronic signal that is processed by the electronic circuit.

In this thesis, we use different scanners based on scintillation detectors (see chapter 2), and the simulation tool developed in chapter 5 is based on this kind of detectors.

1.2.1 Scintillation crystal

In a scintillation material, any gamma photon interaction results in an energy-proportional emission of photons in the visual energy range. Currently, there are different kind of scintillation crystals (organic and inorganic in solid, liquid and gaseous form). In PET scanners, we commonly find solid inorganic crystals because of their high stopping power (defined as the energy lost per unit length) and high light response [Knoll, 2011].

The intrinsic efficiency of the scintillator is a function of the effective atomic number (averaged over its composition), the density, and the crystal thickness. The energy

	NaI	**BGO**	**GSO**	**LSO**	**LYSO**	**LuAP**
Density (g/cm^3)	6.7	7.1	6.7	7.4	7.1	8.3
Z_{eff}	51	74	59	66	60	65
Attenuation coefficient (cm^{-1})	0.34	0.92	0.62	0.87	0.86	0.90
Refractive index	1.85	2.15	1.85	1.82	1.81	1.95
Light yield (%NaI)	100	15	41	75	80	16
Decay constant (ns)	230	300	56	40	41	18

Table 1.2: Main properties of common PET scintillators [Lewellen, 2008].

resolution is closely related to the photons emitted and its transport through the crystal. The number of photons emitted is quantitatively measured as the light yield, and the transport of photons is dominated by the material refractive index (the lowest, the better). The emission spectrum of the scintillator should match the photosensor sensitivity for optimal performance. It is also important to cover the crystal with reflectors to guarantee that a high fraction of the light can be collected. The time response in the scintillator has two main components: a fast rising signal and a slow falling signal (decay constant) that determines the time resolution [Ljungberg et al., 2012]. In the PET detectors, the scintillation crystals are usually pixellated to improve the spatial resolution of the photon interaction point. Some manufacturers have developed monolithic scintillators too [Freire et al., 2021, Stockhoff et al., 2021].

In table 1.2 we have summarized the main charasteristic of the most common scintillators employed in PET.

1.2.2 Photosensors

The photosensor converts the optical response of the scintillator to an electrical signal that is later processed. To perform energy spectroscopy of the incident gamma rays, this device must maintain the proportionality between the light generated and the electrical pulse. Spatially sensitive photosensors are also required to provide the gamma interaction point in the detector. Another quality required for the photosensors in recent multimodal PET/MRI scanners is the insensitivity under magnetic fields.

We summarize here the two most common photosensors employed in PET and their main characteristics:

- **Photomultiplier tubes**. The photomultiplier tubes (PMTs) are made of a vacuum tube that enclosures three main parts: the photocathode, a sequence of dynodes, and the anode [Knoll, 2011]. When the photons hit the photocathode, an electron response is generated. An operation voltage (1kV to 2kV) is applied between the cathode and the anode, stepped through the sequence of anodes with a resistive voltage divider. Thus, we can conduct the electrons from the cathode to the anode, amplifying the signal in every dynode. As a result, the final signal is proportional to the energy deposited in the scintillator. The usual gain of a PMT (electrons in the anode per every electron generated in the cathode) may vary from 10^5 to 10^8. To obtain position sensitive PMTs (PS-PMTs), an array of anodes can be used, generating multiple signals that can be processed to obtain the emission point in the cathode with the Anger logic [Anger, 1969].

  The PMT has been the reference photosensor for PET scanners since the beginning of the technique, but its disadvantages are encouraging the use of other photosensors in the field. Some of these disadvantages are its sensitivity to magnetic fields [Vaquero et al., 2013], and its expensive cost because of complex manufacturing process.

- **Silicon photomultipliers**. The silicon photomultipliers (SiPMs) are conformed of a matrix of avalanche photodiodes (APDs). The APDs are semiconductor detectors that create an avalanche of hole-electron pairs when a visible or ultraviolet photon is absorbed in a depleted region. The APD itself can be used as a photosensor, but they suffer of high noise in the amplification process. For this reason, in SiPMs, the matrix of APDs works on the *Geiger-mode*, in which they detect single photons when a signal is generated over a threshold [McIntyre, 1985]. The multiple-cells can provide a proportional signal to the input light by summing all the activated cells.

  The attraction of SiPMs relies on their gain (similar to PMTs), fast response, low manufacture expenses, compatibility with MRI scanners, and possible one-to-one coupling with the scintillation pixels [Slomka et al., 2016]. They also have some drawbacks, such as non-linear response or lower photon detection efficiency.

1.2.3 Electronics

Once the pulse is generated, the electronic circuit behind the photosensor has to process it. From the pulse, we expect to obtain information of the time stamp, deposited energy, and location of the gamma photon interaction [Knoll, 2011]. The information of the pulse is what we call a single event. The time stamp can be obtained analogically with the help of a constant fractor discriminator (CFD) that estimates the moment in which it exceeds a constant fraction of its maximum amplitude. The total charge accumulation in the photosensor is used to estimate the energy of the photon. Usually there are low- and high-energy discriminators to discard events far from the 511 keV peak. The gamma interaction point is estimated from the multiple outputs of the photosensor.

Recalling the interest of detecting pairs of annihilation photons, the time information of every single event is fed to a coincidence circuit to generate pairs of events within a time window of the order of a few nanoseconds. This is what we call electronic collimation, and we call coincidence event to every pair of single events within the time window. The time difference between both singles, known as time-of-flight (TOF), is very useful to improve the decay location within the LOR [Conti, 2011].

1.3 PET acquisitions

In the previous section 1.2 we have described the PET detectors and the process up to the generation of single and coincidence events. In summary, a single is any signal generated in a detector by the gamma photon interaction, and a coincidence is any couple of singles that are detected within a narrow time window. In this section we are going to classify the coincidence events given their physical origin, and we will give a brief overview of the common data formats employed to organise the data.

1.3.1 Classification of coincidence events

In practice, the coincidences do not always come from an annihilation event. We resume here the different kind of coincidences (figure 1.6).

- **Prompt coincidence**. This is the term employed for any coincidence detected within the defined time window, regardless of its physical origin (any of the examples in figure 1.6).

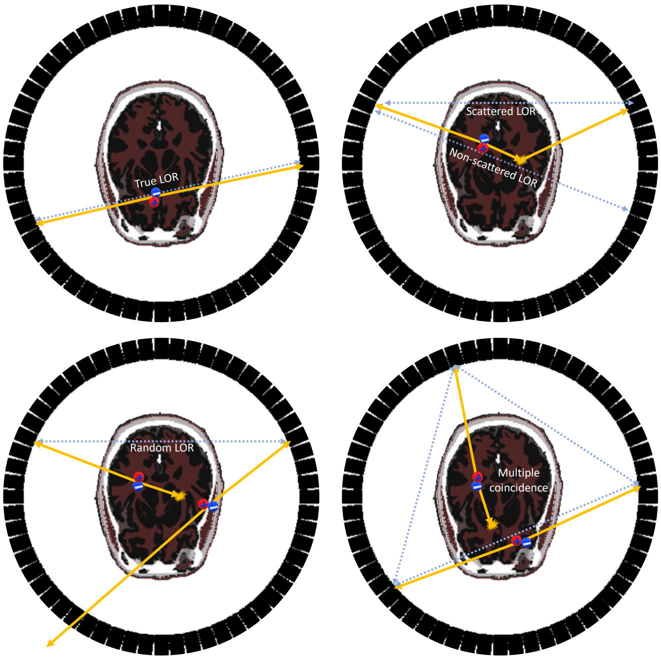


Figure 1.6: Graphical examples of the different coincidences. True (top left), Compton (top right), random (bottom left) and multiple coincidence (bottom right).

- **True coinidence**. When the two singles actually come from the same positron annihilation without suffering any interaction with the patient body, we call it a true coincidence. These events are the ones that provide reliable information about the activity distribution inside the scanner (top left in figure 1.6).

- **Scatter coincidence**. If any of the photons suffers a scatter interaction with the patient, the photon changes its energy and direction as stated in section 1.1.2. When the scattered photon still has sufficient energy to be detected in the energy window, the LOR obtained does not spatially correlate with the nuclear decay. Correction of this events is discussed in section 1.4.3.4, and an approach is experimentally tested in section 5.4.

- **Random coincidence**. Since the nuclear decay is a stochastic process, we may find random coincidence in any case that two photons that do not come from the same annihilation are detected in coincidence (figure 1.6). These events will be more common as we increase the activity inside the scanner field of view. In section 1.4.3.5 we give a further discussion on how to correct these events.

- **Multiple coincidences**. In close relation with the random coincidence, we may find more than two singles within the same time window. In standard PET we simply discard these coincidences to avoid wrong LOR assignments. Some authors have benefited of these coincidence effects to reduce noise or to differentiate multi-tracer acquisitions [Cal-González et al., 2015a, Lage et al., 2015, Lopez-Montes, 2021].

1.3.2 Data organization

In a typical PET acquisition millions of coincidences are measured. In this section we revise different strategies to organize the data.

1.3.2.1 List mode data

This is the most informative format. In list mode data the information from every coincidence is stored. Every coincidence may include the detectors and crystal pixels involved, the absolute time stamp, TOF, the energies deposited by each gamma, and any other detail that the manufacture has considered (in general, list-mode data is scanner-specific). In some cases it is possible to find lists of the single events. This format is very useful to perform dynamic studies as data can be arranged in frames of specific duration work with short frames of data [Álvarez-Gómez et al., 2021, Arias-Valcayo et al., 2019, 2020].

1.3.2.2 LOR histogram

The LOR histogram simply consists of a histogram of the coincidences accumulated in every LOR. Data stored can be considerably reduced at the expense of losing additional information like the event timing or the deposited energy. Provided the histogram format, the reconstructor may recover the physical coordinates of the LORs to generate the image [Kadrmas, 2004]. This is the data format utilized in the reconstruction code detailed in chapter 3.

1.3.2.3 Sinogram

The sinogram is a specific case of the LOR histogram organization based on grouping parallel LORs at fixed angles [Michel Defrise and Paul Kinahan, 1998] (see figure 1.7

in next section 1.4.1). This organization is intuitive since the LORs can be viewed as image *projections* in a fixed linear path through the image. In this format, the LOR is parametrized by the orthogonal distance to the center of the scanner, known as the radial coordinate, and the angular inclination of the LOR in the transverse plane. Separately, the axial distance from the LOR center to the center of the scanner (*slice*) and the *inclination* in terms of the axial difference between the two LOR crystals are given. The slices with zero axial difference are called *direct* sinograms, and the rest of them are called *oblique* sinograms, and each group of sinograms with similar axial difference is called a *segment*. For a given slice and inclination (either for oblique or direct sinograms), the radial and angular coordinates define sinusoidal shapes that name the format (figure 1.7). The visual representation of the data and its connection with the Radon transform (see section 1.4.1) is an advantage over other formats.

Motivated by the cylindrical morphology in common PET scanners [Delso et al., 2011, Wang et al., 2006], the sinograms are further grouped to reduce data storage. First, the users may discard too oblique segments because of their poor resolution quality (because of depth-of-interaction effect, see section 1.4.2.1). The term *maximum ring difference (MRD)* refers to the maximum axial difference allowed, measured in number of detector crystals. Secondly, the axial sinograms with close axial difference and equal axial position are combined to reduce the sinogram storage. The concept of *span* refers to the number of sinograms smashed together. In general notation, the span has to be an odd number higher than 1, and it is equal to the sum of added planes in odd and even planes [Michel Defrise and Paul Kinahan, 1998].

In section 5.4, reconstructed data from a clinical scanner in the sinogram format are shown.

1.3.2.4 Axial rebinning

Sometimes it is required to reduce the sinogram size to decrease reconstruction time, or making the data accessible to 2D reconstruction techniques (see section 1.4.1). Axial rebinning techniques reduce 3D PET sinograms to 2D segments. A fast and simple rebinning method is the Single-Slice Rebinning (SSRB) algorithm [Daube-Witherspoon and Muehllehner, 1987]. SSRB approximates the average axial position of each oblique sinogram and places the coincidences to the closest direct sinogram. A more accurate method is the Fourier rebinning (FORE) [Defrise et al., 1997]. FORE uses the Fourier transform of direct and oblique sinograms to estimate the direct sinograms based on the frequency-distance relationship [Weishi Xia et al., 1995]. Few other variations of this method have been presented [Ben Bouallegue et al., 2007, Xuan Liu et al., 1999]. Recently, our group also presented the pseudoinverse approach for both axial rebinning and 2D reconstruction, achieving similar accuracy to FORE with execution times similar to SSRB [Lopez-Montes, 2021, Lopez-Montes et al., 2017, López-Montes et al., 2020, Lopez Montes et al., 2019]. Even though rebinning methods played an active role in the past, nowadays fully 3D reconstruction algorithms are preferred as its use in current computers is feasible for its daily use.

1.4 PET image reconstruction

The goal of Emission Computed Tomography (ECT), either PET or SPECT, is to obtain the most accurate representation of the radiotracer distribution inside the patient. Different methods have been proposed over the years in order to translate the emission detected events into the 3D image space. Most authors have separated reconstruction methods into analytical and iterative (or statistic) methods [Cherry et al., 2012, Iriarte et al., 2016, Qi and Leahy, 2006, Tong et al., 2010], although it is worth mentioning the recent incorporation of Deep Learning (DL) techniques either for direct reconstruction or mixed reconstruction with existing methods [Arabi et al., 2021, Gong et al., 2021, Reader et al., 2021, Reader and Schramm, 2021].

In order to achieve quantitative image values in ECT, it is of great importance to account for different correction factors during the reconstruction process [Tong et al., 2010, Zaidi and Karakatsanis, 2018]. Some of the most relevant ones are radioactive decay, normalization, attenuation correction, scatter events, random coincidences, dead time, or pile up. These corrections can be applied in advance to the measured data, or incorporated in the reconstruction algorithms.

In this section we will give a brief explanation of the main reconstruction algorithms and PET corrections.

1.4.1 Analytical algorithms

A fundamental tool of these methods is the Radon transform. The Radon transform mathematically describes the 2D projections of an object $f(x, y)$ at a given angle θ and

Figure 1.7: Pictorial scheme of the Radom transform. The coordinates (ρ, θ) in the sinogram domain (right panel) are correlated with the projection at a fixed angle and radial distance in the image domain (left panel).

radial distance to the center ρ [Easton, 2010],

$$p(\rho, \theta) = \int_{D_{(\rho,\theta)}} f(x, y) dl_{(\rho,\theta)}, \tag{1.11}$$

where $D_{(\rho,\theta)}$ is the length of the linear integral path, and $dl_{(\rho,\theta)}$ the differential unit length through this path. The Radon transform defines the projection space, where the PET data are found, also called sinogram in PET (see figure 1.7). To recover the image from the projection space, the central-section theorem plays a mayor role (also called Fourier slice theorem). This theorem states that the Fourier transform (FT) of the 1D projection at fixed angle θ_0 is equal to the projection through the center and at the same angle of the FT of the image [Cherry et al., 2012],

$$\mathcal{F}\{p(\rho, \theta_0)\}(\omega_\rho, \theta_0) = \mathcal{F}\{f(x, y)\}(k_x, k_y)|_{\theta=\theta_0}, \tag{1.12}$$

where ω_ρ, k_x, and k_y are the conjugates of ρ, x and y respectively in the Fourier space. From this theorem we can find a reconstruction method if we have the complete access to the data $p(\rho, \theta)$. The most widely used method is the filtered-backprojection (FBP). The general expression of the FBP is given by [Cherry et al., 2012, Easton, 2010, Tong et al., 2010]

$$f(x, y) = \int_0^\pi \mathcal{F}^{-1}\{h(\omega_\rho)\mathcal{F}\{p(\rho, \theta)\}(\omega_\rho, \theta)\} d\theta. \tag{1.13}$$

Note here that we extend the angular integral only from 0 to π, since PET data are usually ordered with positive and negative values of ρ. $h(\omega_\rho)$ is a filter function that can be modified to regularize the image (the strict FBP equation simply considers $h(\omega_\rho) = |\omega_\rho|$).

Other analytical methods have been developed, such as the backprojection-filtering (BPF) [Alessio and Kinahan, 2006], or three dimensional adaptations of the FBP [Kinahan and Rogers, 1989]. Recently, we have shown how a reconstruction method based on a more accurate model of the system response matrix (see section 1.4.2.1 for further details about the SRM) using a pseudoinverse transformation provides higher resolution than the traditional FBP [López-Montes et al., 2020, Lopez Montes et al., 2019].

In general, analytical methods assume idealized image projection models of the system, and noise-free and complete data that can be mathematically inverted to recover the activity distribution. These assumptions are not good approximations for PET scanners, and it may result in artifacts in the image. The physical effects that degrade the image quality are not modeled in the FBP, thus limiting the resolution that can be recovered from this method. In [Herraiz et al., 2006b], the sinograms were deconvoluted to restore the lost resolution because of intrinsic PET effects. The problem as it has been stated so far is ill-posed, and small changes in the data (for example, because of Poisson noise, random coincidences, scatter events, etc.) may lead to big changes in the image. Equation 1.13 includes a filter $h(\omega_\rho)$ that removees high-frequency noise, thus reducing the components at high values of ω_ρ. A few common examples are the ramp filter, the Hamming filter or the Shepp-Logan filter [Cherry et al., 2012]. PET scanners are also conformed of discrete detector units with gaps in the projection domain from two contiguous blocks, that result in discrete data with small gaps. This problem is partially solved using "inpainting" techniques to fill in the empty spaces, or other sophisticated strategies [Herraiz et al., 2008]. Most of the analytical methods are derived from 2D formulation, whereas PET data are gathered in 3D. Although some authors adapted the problem to 3D [Kinahan and Rogers, 1989], it is very common to use 2D rebinning (see section 1.3.2.4) techniques followed by 2D reconstruction [Ben Bouallegue et al., 2007, Daube-Witherspoon and Muehllehner, 1987, Defrise et al., 1997, Lopez-Montes et al., 2017, López-Montes et al., 2020, Lopez Montes et al., 2019, Xuan Liu et al., 1999]. Analytical methods produce noisier images than iterative methods (see next section 1.4.2). In contrast, it is important to remark that analytical methods are usually faster and being a linear mapping between the data and the reconstructed images, they are typically preferred for performance assessment of PET scanners [National Electrical Manufacturers Association, 2008].

1.4.2 Iterative algorithms

Iterative methods are based on the optimization of an objective function, searching for the image that better matches the data through small steps. These algorithms include a system response matrix that relates the data space with the image domain, the objective function to be optimized, a statistical model of the noise in the data, and an iterative algorithm that estimates the image evolution toward the convergence of the optimal solution [Alessio and Kinahan, 2006, Michel Defrise and Paul Kinahan, 1998].

In practice, iterative methods can be viewed as comparison and update algorithms [Cherry et al., 2012] (see figure 1.8). The process starts with an arbitrary initial image, usually a uniform distribution filling the whole field of view of the scanner. In each iteration the current image is forward projected to obtain a projection estimate of the data, and we compare it with the measured data to obtain the correction factors given by our objective function. The image correction is computed with a backprojection operator that takes the correction factors back from the projection domain to the image domain. The optimization algorithms apply these corrections to update the image. In this process, we expect to reduce the differences between our model and the data until convergence condition is achieved.

One of the keys of iterative reconstruction is the forward projection step. We can describe this transformation in matrix representation as

$$\tilde{Y} = AX + R + S, \tag{1.14}$$

where $\tilde{Y} \in \mathbb{R}^N$ is an array of length N representing the expected values in the projection space, $X \in \mathbb{R}^M$ is the reconstructed object of M voxels, and $A \in \mathbb{R}^{N \times M}$ is the system response matrix (SRM) that relates the probability of having a detected coincidence in a projection unit $\tilde{y}_i$ from a given voxel x_j (i.e. each matrix component a_{ij}). $R, S \in \mathbb{R}^N$ correspond to additional effects, such as random coincidences or scatter events (respectively) that are usually subtracted from the data in analytical algorithms. The SRM is of great importance to obtain the best image quality of our system, and it

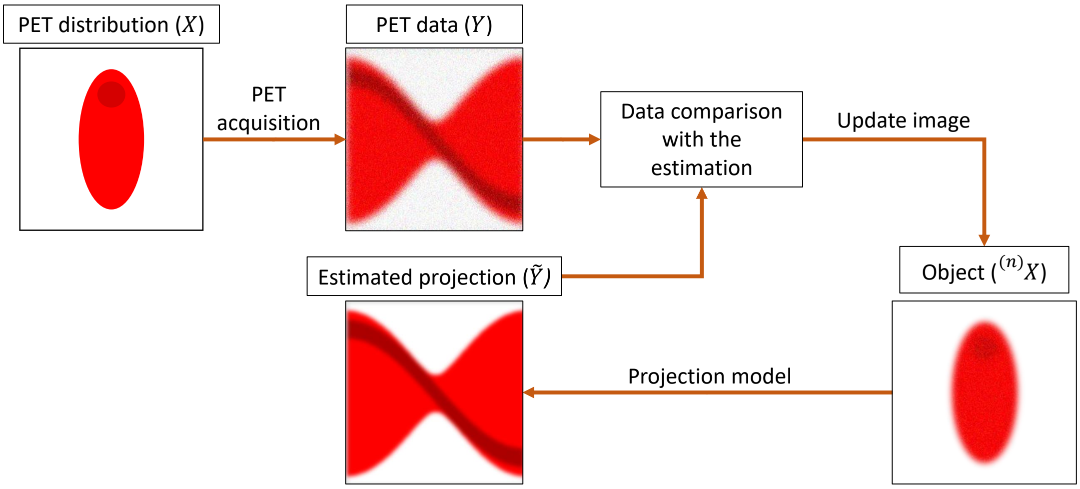


Figure 1.8: Graphical workflow of iterative reconstruction algorithms. The noisy data acquired from the acquisition (Y) of a PET distribution (X) are iteratively compared against an estimate projection ($\tilde{Y}$), until the object ($^{(n)}X$) has converged.

must accurately model the actual physics in our acquisition system [Iriarte et al., 2016]. We describe in detail the different models proposed for the SRM in section 1.4.2.1.

The objective function also plays an important role in the final solution achieved. There are a set of algorithms, known as Algebraic Reconstruction Techniques (ART) [Gordon et al., 1970], that minimize the least squares problem of the system ($\sum_i (y_i - \tilde{y}_i)^2$). Later, a variation of the ART algorithm was developed by Andersen [1984], the Simultaneous Algebraic Reconstruction Technique (SART). One of the most successful algorithms during the last forty years is the Maximum Likelihood Expectation Maximization (MLEM) [Shepp and Vardi, 1982]. Motivated by the nuclear nature of PET and SPECT, the objective function in this case is the log-likelihood of the Poisson model for the data and the projections, and the optimizer chosen is the expectation maximization [Dempster et al., 1977]. In Browne and de Pierro [1996], the poissonian log-likelihood is optimized via the gradient ascent algorithm, and the MLEM is presented as a particular case of this framework. The Image Space Reconstruction Algorithm (ISRA) [Daube-Witherspoon and Muehllehner, 1986] was presented to preserve non-negativity images and reduce the memory requirement of the MLEM. In Reader et al. [2011], both ISRA and MLEM are presented from the point of view of a gradient descent algorithm of a weighted least-squares objective function, with properly chosen descent steps. In terms of noise models, the weighted least squares can be viewed as the log-likelihood of a gaussian model of the noise. In general, iterative methods converge to noisy images if the number of iterations is not controlled. To avoid this, penalty functions (also called priors) are added to the objective function [De Pierro, 1995, De Pierro and Yamagishi, 2001, Green, 1990]. The goal is to select the image that satisfies our expectations, preserving edges while imposing local smoothness [Alenius and Ruotsalainen, 1997, Bettinardi et al., 2002, Nuyts et al., 2002, Sidky and Pan, 2008]. In section 1.4.2.2 we will give more details about the MLEM algorithm, and in section 1.4.2.4 we will talk about Bayesian penalized likelihood methods.

All of these components make iterative algorithms outperform analytical algorithms [Qi and Leahy, 2006] in terms of signal to noise ratio. The statistical model makes them more robust to noise and incomplete data, and together with the iterative scheme allow for non-linear considerations in the reconstruction workflow. They can also introduce geometrical properties and model physical processes in the SRM, thus making them more accurate and easily adaptable for 3D imaging [Tong et al., 2010]. Iterative methods are computationally intensive compared with the FBP, since each iteration requires a forward projection and a backward projection operation, whereas in FBP this step is performed only once. Nevertheless, nowadays computer power and GPU parallel computing allow for routine image reconstruction with iterative methods in short computation time [Barker et al., 2009, Galve et al., 2020b, Herraiz et al., 2011, Meng et al., 2019]. Furthermore, ordered subsets algorithms [Ahn and Fessler, 2003, Hudson and Larkin, 1994] reduce the reconstruction time by the use of a small set of data in each update. The Ordered Subsets Expectation Maximization (OSEM) method is reviewed in section 1.4.2.3.

1.4.2.1 The System Response Matrix

In equation 1.14 we have presented the System Response Matrix (SRM) A as a matrix connecting the image and projection space. The accurate correspondence of the SRM with the actual response of the scanner is crucial to achieve optimal image quality [Iriarte et al., 2016]. The most important considerations are the scanner geometry, the positron range, the non-collinearity of annihilation photons, photon interactions with the patient body (attenuation and scatter), random coincidences, and the detector response (such as the depth of interaction in the detectors (DOI), inter-crystal scatter, or intrinsic

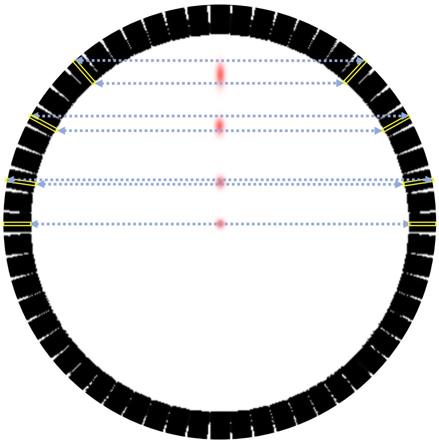


Figure 1.9: Visual representation of a point source resolution loss in the radial direction caused by the DOI effect when we approach the edge of the scanner.

efficiency). The DOI effect [Mohammadi et al., 2019], which has not been explained so far, is caused by the uncertainty in the point of interaction of the gamma ray within the crystal thickness (figure 1.9). Usually, the matrix is factorized and attenuation correction, scatter events, random coincidences, and the normalization (sensitivity of a given LOR) are corrected separately [Comtat et al., 2004, Michel et al., 2000, Zaidi and Karakatsanis, 2018]. For this reason, we will talk in more detail about these factors in section 1.4.3.

The SRM has been estimated using either analytical, experimental, Monte Carlo, or hybrid approaches. Analytical approaches are based on geometrical projection models [Kang et al., 2021, Lougovski et al., 2014, Markiewicz et al., 2014, Siddon, 1985, Zeng et al., 1991], analytical models of the physics [Huesman et al., 2000, Li et al., 2015, Moehrs et al., 2008, Pratx et al., 2009], or convolutional kernels usually referred to as Point Spread Function (PSF) [Reader et al., 2003]. The PSF is a widely used method for SRM modeling [Rogasch et al., 2021], consisting of either image or projection blurring kernels applied as a convolution operation. These methods are fast, and they allow for on-the-fly calculations that drastically reduce the storage requirements of a precomputed SRM [Gillam and Rafecas, 2016, Iriarte et al., 2016]. For this reason, they have been usually combined with experimental or MC estimations of the SRM to optimize the computational performance. In experimental approaches, point sources are acquired at different positions of the scanner field of view and the measured response is fitted to hand-crafted convolutional kernels either in the image [Kotasidis et al., 2011, Miranda et al., 2020, Rapisarda et al., 2010] or sinogram domain [Alessio et al., 2005, Gong et al., 2017, Panin et al., 2006, Tohme and Qi, 2009, Zhou and Qi, 2011]. In [España et al., 2007], the authors used narrower PSF for single crystal events identified by the detectors to increase image resolution. Other authors have benefited of MC methods to reproduce the SRM with long simulations. In this case, the SRM may be determined as an explicit matrix of probabilities relating each voxel-LOR pair [Chien-Min Kao et al., 2007, Meng et al., 2019, Rafecas et al., 2004, Southekal et al., 2011, Zhang et al., 2010]. Since these matrices require high loads of memory storage, symmetries and SRM sparsity have been exploited to reduce the SRM size. To avoid size limitations, some authors adapted the simulation results to more complex descriptions, such as PSF descriptions [Li et al., 2015, Pilleri et al., 2019, Wei and Vaska, 2020, Xu et al., 2019, Zhou and Qi, 2014], or wide Tubes of Response (TORs) [Herraiz et al., 2006a, Meng et al., 2021] instead of simple linear projection models. Other approaches include separate positron range correction, that is included in the reconstruction as an image blurring kernel [Cal-Gonzalez et al., 2009, Cal-González et al., 2015b, Kraus et al., 2012].

1.4.2.2 Maximum Likelihood Expectation Maximization (MLEM)

As it was aforementioned above, the Maximum Likelihood Expectation Maximization algorithm (MLEM) [Shepp and Vardi, 1982] arises from the maximum likelihood with a Poisson model of the data, together with the Expectation Maximization (EM) algorithm [Dempster et al., 1977]. From the Poisson distribution, we can state the probability of having y_i coincidences in a LOR i, given an image X, as

$$P(y_i|X) = \frac{e^{-\tilde{y}_i}\tilde{y}_i^{y_i}}{y_i!}, \tag{1.15}$$

where y_i is each component of a vector $Y \in \mathbb{R}^N$, representing the number of coincidences in a LOR i, and $\tilde{y}_i$ is the expected value for y_i, as defined in equation 1.14.

The likelihood function, is thus defined by the multiplication of all the individual Poisson contributions,

$$L(Y|X) = \prod_i P(y_i|X) = \prod_i \frac{e^{-\tilde{y}_i}\tilde{y}_i^{y_i}}{y_i!}. \tag{1.16}$$

This function is hard to manage, and the log-likelihood (equation 1.17) is preferred for the optimization,

$$l(Y|X) = log(L(Y|X)) = \sum_i -\tilde{y}_i + y_i ln(\tilde{y}_i) - ln(y_i!). \tag{1.17}$$

In terms of maximization, both the likelihood and the log-likelihood are concave functions (as it was proved in [Shepp and Vardi, 1982]), and the optimization of either one of them will converge to a unique solution. The log-likelihood in 1.17 can be maximized applying the Kuhn-Tucker conditions, which are the following for the proposed problem:

$$\begin{aligned} &1) \quad x_j \left.\frac{\partial l(Y|X)}{\partial x_j}\right|_X = 0 \\ &2) \quad \left.\frac{\partial l(Y|X)}{\partial x_j}\right|_X \leq 0 \quad \text{if } x_j = 0, \end{aligned} \tag{1.18}$$

where x_j is each of the components of the image X. The derivative of equation 1.17 results in

$$\frac{\partial l(Y|X)}{\partial x_j} = -\sum_i a_{ij} + \sum_i a_{ij} \frac{y_i}{\tilde{y}_i}. \tag{1.19}$$

Recall that a_{ij} refers to the matrix element of the SRM A in 1.14. Finally, if we introduce equation 1.19 in the first condition of 1.18, using a fixed-point optimization procedure we can obtain the following iterative solution for X:

$$x_j^{(n+1)} = \frac{x_j^{(n)}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\tilde{y}_i^{(n)}}, \tag{1.20}$$

where $\tilde{y}_i^{(n)}$ is the estimate of y_i given the image $X^{(n)}$ at an iteration n. In [Shepp and Vardi, 1982] this solution is understood as an EM algorithm: 1) the expectation step of the unknown $\tilde{y}_j^{(n)}$ (the total emitted data from a voxel j given the uncomplete data Y and the image $X^{(n)}$) correspondent to the right hand side of equation 1.20, and 2) the maximization step of the log-likelihood (update of $x_j^{(n+1)} = \tilde{y}_j^{(n)}$).

The properties of equation 1.20 make the MLEM very attractive for image reconstruction [Qi and Leahy, 2006]. The non-negativity of X is guaranteed, the log-likelihood is monotonically increased, i.e. $l(Y|X^{(n+1)}) > l(Y|X^{(n)})$, and the count preservation (at any iteration (n) we find that $\sum_i \tilde{y}_i^{(n)} = \sum_i y_i$). Its pitfalls arise for high number of iterations, when the noise starts to dominate the image, besides its pure form is computationally intensive. For this reasons, the MLEM algorithm has been widely studied for the last forty years, trying to improve its performance in a wide variety of related works [Ahn and Fessler, 2003, Alessio and Kinahan, 2006, Browne and de Pierro, 1996, De Pierro, 1995, Deidda et al., 2019, Green, 1990, Herraiz et al., 2006a, Hudson and Larkin, 1994, Iriarte et al., 2016, Qi and Leahy, 2006, Reader et al., 2021, Tong et al., 2010, Wang and Qi, 2015, Zaidi and Karakatsanis, 2018]. In the following sections we will give a deeper insight on two variations of the method, the Ordered Subsets Expectation Maximization (section 1.4.2.3), and Bayesian penalized likelihood methods (section 1.4.2.4).

1.4.2.3 Ordered Subsets Expectation Maximization (OSEM)

PET data acquired in modern scanners can be composed of up to a thousand billions of LORs [Delso et al., 2011, Udias et al., 2018, Wang et al., 2006], thus making the forward projection in equation 1.20 (computation of $\tilde{y}_i$) a huge computational step that enlarges the reconstruction process. Furthermore, adequate image convergence may require up to a hundred updates.

To overcome this issue, Hudson and Larkin [1994] derived the Ordered Subsets Expectation Maximization algorithm (OSEM). The summation over the whole set of LORs N is divided into several subsets K, thus increasing the number of updates for a complete projection step. The equation 1.20 in the OSEM algorithm reads as follows:

$$x_j^{(n,k+1)} = \frac{x_j^{(n,k)}}{\sum_{i \in S_k} a_{ij}} \sum_{i \in S_k} a_{ij} \frac{y_i}{\tilde{y}_i^{(n,k)}}, \tag{1.21}$$

where S_k refers to the LORs within a subset k. The convergence speed is proportional to the number of updates, and therefore the computation time is reduced by the subsets number ($1/K$) [Alessio and Kinahan, 2006].

The use of subsets can be generalized to other optimizers, like Browne and de Pierro [1996] did for the ascent gradient algorithm, later combined with a penalized likelihood prior [Ahn and Fessler, 2003, De Pierro and Yamagishi, 2001].

1.4.2.4 Bayesian Penalized Likelihood methods

Due to statistical noise in the data, even if we consider the Poisson model in the objective function, the reconstruction problem is ill-conditioned and iterative methods find a trade off between noise and resolution for a high number of iterations. The number of iterations might be used as an arbitrary stopping criteria to avoid noisy images, among other heuristic methods [Guo and Renaut, 2011].

In order to limit noise in the reconstructed images, different regularization methods based on Bayesian penalized likelihood have been developed [Qi and Leahy, 2006, Reader et al., 2021]. *A priori* information of the image is considered to compute the posterior of the image derived from the Bayes' theorem [Ahn and Fessler, 2003, De Pierro, 1995, De Pierro and Yamagishi, 2001, Green, 1990, Mumcuoglu et al., 1996],

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \tag{1.22}$$

Taking logarithms and discarding constants, we can induce the new objective function,

$$\Phi(X) = l(Y|X) + log(P(X)), \tag{1.23}$$

where $\Phi(X)$ is the new objective function, $l(Y|X)$ is the log-likelihood as it was described in equation 1.17, and $P(X)$ is the penalty function or prior. These methods are also called maximum *a posteriori* (MAP), since they estimate the posterior of the image. The penalty function $P(X)$ is manually tailored based on assumptions on the image. Many authors have simplified this equation considering Gibbs priors of the form $P(X) \propto exp(-\beta Q(X))$, β being a control parameter and $Q(X)$ being a function of the differences among voxel values within a local neighborhood,

$$Q(X) = \sum_j \sum_{j' \in N_j} \psi(x_j - x_{j'}). \tag{1.24}$$

Examples for $\psi(x)$ are the quadratic function or the Huber functions [Mumcuoglu et al., 1996], the Geman function [Nuyts et al., 2002], the relative differences prior (RDP) [Ahn et al., 2015, Asma et al., 2012, Nuyts et al., 2002], the median root prior (MRP) [Alenius and Ruotsalainen, 1997, Alenius et al., 1999, Bettinardi et al., 2002], or patch based priors [Guobao Wang and Jinyi Qi, 2012]. In general, these methods are susceptible to proper hyper parameter selection (arbitrary value of β). DL techniques may overcome this issue using more sophisticated penalty functions [Reader et al., 2021].

Equation 1.23 with the Gibbs prior defined in 1.24 can be optimized using the one step late or maximum *a posteriori* (MAP) method [Green, 1990], so called MAP-EM,

$$x_j^{(n,k+1)} = \frac{x_j^{(n,k)}}{\sum_{i \in S_k} a_{ij} + \beta \left. \frac{\partial Q}{\partial x_j} \right|_{x_j^{(n,k)}}} \sum_{i \in S_k} a_{ij} \frac{y_i}{\tilde{y}_i^{(n,k)}}. \tag{1.25}$$

1.4.3 Data corrections

We already discussed in section 1.4.2.1 about the SRM relevance to model the physics behind the positron annihilation and the scanner response, but there are still a few corrections left that are usually accounted separately for proper quantitative and artifact-free imaging. In this section we will summarize the most common ones, and we will give a description on the typical methods applied to mitigate them.

1.4.3.1 Decay

The nuclear decay of the radiotracer has to be taken into account for accurate quantification studies [Zaidi and Karakatsanis, 2018]. This is even more important in dynamic studies. We can correct this factor from the exponential nuclear decay law. The number of counts N_i in the frame i of and acquisition is given by

$$N_i = \Gamma_{e^+} \cdot \int_{t_i}^{t_i + \Delta t_i} A_0 \cdot e^{-\lambda t} dt = \Gamma_{e^+} \frac{A_0}{\lambda} e^{-\lambda t_i} \left(1 - e^{-\lambda \Delta t_i}\right), \tag{1.26}$$

where Γ_{e^+} is the β^+ branching ratio, λ is the decay constant of the radionuclide, and t_i and Δt_i are the initial time and time length in frame i.

To ensure that we reconstruct similar average activity ($A_i = N_i / \Delta t_i$) for all the frames, we should correct for the frame-dependent terms. We can use the following decay factor to obtain similar initial activity $A_0 = A_i D_i$ [Bailey et al., 2005]:

$$D_i = \frac{\Delta t_i \lambda e^{\lambda t_i}}{1 - e^{-\lambda \Delta t_i}}, \tag{1.27}$$

where Γ_{e^+} has been neglected. Note that this is a scale factor that may be applied after the image reconstruction.

1.4.3.2 Normalization

The normalization correction accounts for the relative sensitivity of a given LOR due to inherent factors of the scanner. The goal is to harmonize the acquired data, so that two equivalent LORs measure the same amount of counts when a source equally illuminates both of them. Among the factors affecting a LOR sensitivity, we can mention the intrinsic efficiency of the detectors due to variations in the light guides or the photomultipliers gains, crystal efficiency due to position in a block, geometrical related factors, differences in the energy or timing calibrations, or any other inaccuracy in the scanner manufacturing process [Bailey et al., 2005, Cherry et al., 2012, Theodorakis et al., 2013]. We can account for the normalization as a multiplicative factor in the SRM, or either correct the data measured by the scanner.

The normalization correction factors might be estimated measuring a known source of activity, such as a rotating line source [Defrise et al., 1991], an annulus, or a cylinder [Vicente et al., 2006]. The direct normalization consists of the estimation of the ratio between the measured activity and the known expected activity [Cherry et al., 2012, Theodorakis et al., 2013]. This approach typically requires long acquisitions to get enough data and reduce the noise. To overcome this issue, component-based normalization decompose the normalization in a variety of factors, considering separately the crystal efficiency, geometrical factors, and other defects (such as count rate dependence) [Badawi et al., 1998, Badawi and Marsden, 1999, Defrise et al., 1991, Hoffman et al., 1989, Niu et al., 2012, Rodriguez et al., 2007].

1.4.3.3 Attenuation correction

The attenuation effect is caused by lost coincidences in the LOR because of photon interactions with the patient body or other objects inside the scanner [Bailey et al., 2005, Cherry et al., 2012]. This is one of the larger effects in PET. Fortunately we can correct attenuation correction (AC) if we can estimate the interaction probability of the photons through the patient body [Bailey et al., 2005, Cherry et al., 2012]. In general, AC factors are LOR-dependent, and they can be included in the forward projection model (the SRM) or applied to correct the acquired data, similarly to the normalization. This attenuation factors are given by

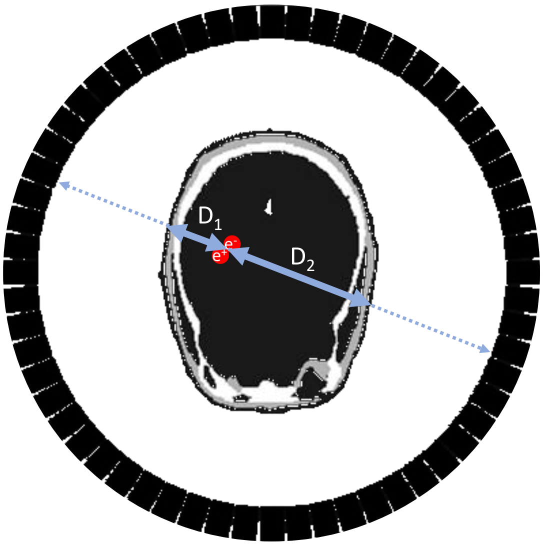


Figure 1.10: Graphical representation of the integrail in equation 1.29

$$u_i = e^{-\int_{D_i} \mu(\vec{r}) dl_i}, \tag{1.28}$$

where u_i is attenuation factor of a given LOR i, D_i is the path through the LOR i, and $\mu(\vec{r})$ is the total attenuation coefficient (considering photoelectric effect, Compton and Rayleigh interactions) inside the patient for 511 keV photons. Note that we consider any scatter interaction as an attenuation, even if the coincidence is detected in a different LOR. Scattered event will be covered in next section 1.4.3.4. It is easy to prove that the final result does not depend on the emission point inside the line D_i, since the interaction probability is given by the multiplication of the attenuation factors of both antiparallel photons (we need to detect both to have a coincidence). Therefore, the two line integrals will be added in the exponential (see fig. 1.10),

$$u_{i,1} u_{i,2} = e^{-\int_{D_{i,1}} \mu(\vec{r}) dl_i} e^{-\int_{D_{i,2}} \mu(\vec{r}) dl_i} = e^{-\int_{D_{i,1}+D_{i,2}} \mu(\vec{r}) dl_i} = u_i. \tag{1.29}$$

The computation of the attenuation factors can be achieved with a transmission scan [Bailey et al., 2005, Cherry et al., 2012]. A rod source of ^{68}Ge-^{68}Ga of equal length to the axial field of view (FOV) of the scanner is rotated around the patient body. A blank scan (without subject) is performed first, to compare with the transmission scan. However, this method provide noisy AC factors [Alenius et al., 1999, Meikle et al., 1993]. The advent of PET/CT and PET/MRI hybrid scanners [Cal-Gonzalez et al., 2018a, Nadig et al., 2021] gives access to anatomical maps that can be related with more accurate attenuation maps (so called µ-map). The µ-map generated is forward projected to obtain the line integral of equation 1.28 and compute the AC factors. In the case of CT images, the Hounsfield Units (HU) can be related to the linear-attenuation coefficients with the use of a bilinear function [Burger et al., 2002, Nakamoto et al., 2002, Schneider et al., 2000] (see equation 5.7). MRI images, in contrast, are not directly correlated to tissue densities, and other methodologies have been explored [Izquierdo-Garcia and Catana, 2016, Ladefoged et al., 2017, Torrado-Carvajal et al., 2019, Wagenknecht et al., 2013]. Time of flight (TOF) information can also be used to obtain the AC, up to a constant factor [Rezaei et al., 2012, 2014]. A bunch of AC correction methods combined with DL techniques are summarized in section 1.4.4.

1.4.3.4 Scatter

The vast majority of the 511 keV attenuated photons in tissue have interacted via Compton scatter, thus losing energy and changing their direction. If the scanner eventually detects a scattered coincidence, it is clear that the assigned LOR to that coincidence will not correspond to the actual one (see figure 1.6). As a result, the accumulation of scatter events results in a smooth background that reduces the accuracy of quantitative measurements [Afshar-Oromieh et al., 2017, Heußer et al., 2017, Lindemann et al., 2019]. This effect will be more prominent in clinical scanners with large patients, where the proportion of scattered coincidences respect to the total ones (i.e. scatter fraction) is up to 50% of the total coincidences [Adam et al., 1999, Ye et al., 2011].

Most PET scanners use an energy window around 511 keV [Delso et al., 2011, Kang et al., 2021, Qi et al., 2011, Samanta et al., 2021, Udias et al., 2018, Wang et al., 2006], approximately from 400 keV to 650 keV, to discard coincidences that may have lost part of their energy in Compton interactions within the patient's body. Nevertheless, many Compton events can still be detected within the photopeak energy window. We can use different methods to estimate the distribution of scatter events. This contribution may be added to the forward projection model (as was stated in equation 1.14) or subtracted to the data. Among the wide variety of methods found in the literature, we can summarize them in experimental, analytical or MC methods [Castiglioni et al., 1999]. Experimental methods rely on the use of different energy windows [Álvarez-Gómez et al., 2021, Castiglioni et al., 1999]. The initial analytical methods assumed a Gaussian distribution for the scatter sinogram, such as one of the methods proposed in [Castiglioni et al., 1999]. This approach lacks of a realistic model of the scattering process. More accurate is the Single Scatter Simulation (SSS) [Ollinger, 1996, Watson, 2000, Watson et al., 1996]. The SSS integrates the single scatter contribution of every LOR-voxel connection to estimate the scatter distribution. The method needs a µ-map to estimate the probability of interaction at every point of the photons track, and a precomputed activity distribution. To compute the global scale of the distribution some authors use tail-fitting techniques [Thielemans et al., 2007, Werling et al., 2002]. This method assumes that all the data in the LORs outside the patient body are caused by the scattered photons. Thus, the scatter distribution is scaled to match the sinogram data in these regions. However, tail-fitting may raise different issues, such as halo artifacts around high organ-to-background regions (see figure 5.14 in section 5.4) [Heußer et al., 2017, Lindemann et al., 2019], and short MC simulations can be computed for this purpose [Pönisch et al., 2003, Ye et al., 2014]. Pure MC simulations are considered the most accurate method [Castiglioni et al., 1999, Holdsworth et al., 2001, Kyung Sang Kim et al., 2014, Ma et al., 2020]. Finally, following the recent trend and its outstanding performance, many authors have benefited of DL for scatter correction [Arabi et al., 2020, Berker et al., 2018, Laurrent et al., 2020, Qian et al., 2017, Shiri et al., 2020, Xiang et al., 2020, Yang et al., 2019, 2021]. In section 5.4 of this thesis we use the UMC-PET simulator for fast MC scatter correction.

1.4.3.5 Random coincidences

When a photon generates a single event in one of the scanner detectors, the scanner opens a time window of the order of 1 ns to admit the second photon in coincidence. However, it may happen that within that time window, more than 1 decay occurs, and therefore, photons from two or more positron annihilations may be detected, producing what is called a random coincidence. In that situation, the second photon may not be spatially correlated with the first one (see figure 1.6). Random coincidences generate an homogeneous background in the image that will deteriorate image contrast [Bailey et al., 2005, Cherry et al., 2012]. The proportion of random coincidences will depend on several factors. Large FOV increases sensitivity and the count rate, and therefore, the proportion of random events [Badawi et al., 2000]. Too wide time coincidence window may accept random coincidences without a benefit in the image quality. The use of a high activity increases the count rate, and therefore, it is directly correlated with higher random rates.

For random corrections, the most typical methods are the time delayed coincidence window [Badawi et al., 1999, Byars et al., 2005] and the singles rate method [Oliver and Rafecas, 2008, Stearns et al., 2003]. A time delayed coincidence window detects coincidences of uncorrelated events (if the time delay is much greater than the coincidence window). This method suffers from statistical noise, since the uncorrelated rates will have a Poisson distribution. In the case of the singles rate method, the randoms rate is given by

$$r_i = 2\tau s_{i,1} s_{i,2}, \tag{1.30}$$

where r_i is the randoms rate in LOR i, 2τ is the total coincidence window of the system [Knoll, 2011], and $s_{i,1}$ and $s_{i,2}$ are the singles rate of the two detector units in LOR i. This equation estimates both prompt and random events, and thus prompt rates should be subtracted to calculate the actual randoms rate in every LOR. In any of the mentioned methods, the randoms estimation can be subtracted to the data or added to the projection model (similar to the scatter correction).

1.4.3.6 Dead time

After a photon generates a signal in the detector, each component of the PET scanner processes the signal during a short fraction of time. Any further signal generated during this time is lost. It may be considered that the detector is "dead" for this fraction of time, so called dead time [Knoll, 2011]. Given the stochastic nature of nuclear decay, there is always a finite probability to lose events because of dead time, thus losing the proportionality between the actual activity inside the FOV and the measured count-rate. This ratio is called the fractional dead time, and it strongly depends on the count-rate. The dead time of each subsystem of the scanner can be experimentally measured to apply the correction, or use analytical methods [Vicente et al., 2013] based on models of the scanner that are fitted to decaying experiments.

The most common models of dead time behaviour can be divided into paralyzable and non-paralyzable [Knoll, 2011]. In the paralyzable model, every time an event falls in the dead time window, an additional dead time period is added, thus increasing the fractional dead time. A simple formula representing this model is given by

$$m(n) = ne^{-n\tau}, \tag{1.31}$$

where m is the measured count-rate, n is the actual count-rate in the detectors, and τ is the dead time. The measured count-rate m in systems approximated to this model drops to zero for high count-rates.

In the non-paralyzable model, every photon detected during the dead time is lost, but no further dead period is added. Thus, the measured count-rate asymptotically tends to τ^{-1} for high count-rates. This model can be approximated by the following equation:

$$m(n) = \frac{n}{1 + n\tau}. \tag{1.32}$$

1.4.3.7 Pile up

When a photon deposits part of its energy in a detector while an earlier photon pulse is being integrated, the detector adds up both pulses, making them indistinguishable. We call these events pile-up. If the pile-up event accumulates energy beyond the upper window threshold, the event is lost. The event can also fall within the energy window, but in that case it will be mispositioned because of the contribution of the second photon. In any case, the image quality will be deteriorated if the amount of pile-up events is high [Badawi and Marsden, 1999, Rodriguez et al., 2007, Vicente, 2012, Vicente et al., 2007].

1.4.4 Deep Learning implementation in image reconstruction

Deep Learning (DL) is a field within artificial intelligence (AI) that consists on artificial neural networks, composed of a large number of layers and neurons able to create complex mappings between inputs and outputs. It is usually composed of a sequence of linear convolution kernels and nonlinear operations, known as Convolutional Neural Networks (CNN) [Simonyan and Zisserman, 2014, Skansi, 2018]. Given an arbitrary cost function, the CNN parameters are "trained" (or fitted, in non-AI language) to achieve the optimal performance [Simonyan and Zisserman, 2014, Skansi, 2018]. This process is computationally intensive, and thus it has not been until the recent evolution of parallel computing of graphic processing units and tensor processing units (GPUs and TPUs, see section 1.6) that these techniques have become available for practical use. However, the inference (i.e. the application of the network to "infer" the desired result from new data) is fast, hence very practical. The success of these techniques is determined by the architecture of the CNN, the training data set, the cost function that relates the data and the parameters, and the optimizer used for the training.

In image reconstruction, the application of DL has followed different strategies [Gong et al., 2021, Reader et al., 2021, Reader and Schramm, 2021]. The direct application is based on image post-processing [Cui et al., 2021, 2019, Herraiz et al., 2020, Schramm et al., 2021], working on the image domain to improve image quality or denoising. Other authors have embedded the CNN within the image reconstruction process, combining them for image regularization [Corda-D'Incan et al., 2020a,b, Gong et al., 2019]. The direct reconstruction from projection data has also been implemented [Feng et al., 2021, Häggström et al., 2019, Whiteley et al., 2020, Zhu et al., 2018], outperforming traditional algorithms.

Quantitative PET has also benefited of DL [Arabi et al., 2021]. In Lee [2021], we can find a recent review about current status and challenges of AC and its combination with DL focused on MRI to pseudo-CT transformation [Blanc-Durand et al., 2019, Sari et al., 2021, Torrado-Carvajal et al., 2019] and anatomical-independent DL [Arabi and Zaidi, 2021, Dong et al., 2019, Hwang et al., 2019, Yin and Obi, 2021]. The scatter correction is tightly bounded to the µ-map, and usually it is jointly corrected with the AC estimation in the image domain [Arabi et al., 2020, Shiri et al., 2020, Yang et al., 2019, 2021]. A different approach, similar to traditional methods for scatter estimation, is focused on the DL estimation in the projection domain [Berker et al., 2018, Laurrent et al., 2020, Qian et al., 2017, Xiang et al., 2020]. In motion estimation and correction, DL has also been implemented [Balakrishnan et al., 2019, Li et al., 2020, Morales et al., 2019].

In spite of its current progression, DL techniques are beyond the scope of this thesis, and no implementation of them is studied along the work presented in the following chapters. Nevertheless, it is worth mentioning that some of the accurate and fast tools developed in this thesis (such as the MC code from chapter 5 or the MC-based reconstruction method of chapter 6) can be used to generate training data for DL methods.

1.5 Monte Carlo methods

Monte Carlo (MC) methods are a class of numerical methods that offer a probabilistic solution to complex mathematical problems, usually applied when the analytical solution is not achievable (such as complex integrals). The essence of MC methods is the use of random numbers to sample variables that represent stochastic processes [Kalos and Whitlock, 2008]. For example, the distribution of detected radiation, or the dose left inside a predefined object from a radioactive sample, can be approximated with realistic MC simulations that model simple physical stochastic processes involved in the particle tracks [Ljungberg et al., 2012]. In section 5.1, we show a brief review of the main MC models for particle tracking used in nuclear physics, the main software developed for medical physics, and specific codes adapted for PET simulations.

In this section, we describe the basics of probability in subsection 1.5.1, different algorithms used for random sampling in subsection 1.5.2, and common variance reduction methods to speed up MC simulation in subsection 1.5.3. The methods presented have been implemented in the development of the Ultra-fast Monte Carlo PET simulator described in chapter 5.

1.5.1 Basics of probability

The probability density function (PDF) $p(x)$ is defined as the function that defines the distribution of a random variable X [Agency, 2019, Kalos and Whitlock, 2008, Ljungberg et al., 2012]. The PDF must be positive, and it must be normalized to the unit,

$$\int_{-\infty}^{\infty} p(x)dx = 1. \tag{1.33}$$

The probability of having a sample X is given by $p(x)dx$, and the probability of having a value lower than x is defined as the cumulative density function (CDF),

$$P(x) = \int_{-\infty}^{x} p(x')dx' := Probability\ of\ having\ X < x. \tag{1.34}$$

The CDF is a positive nondecreasing function that satisfies $P(-\infty) = 0$ and $P(\infty) = 1$. From 1.34 we can easily demonstrate that the probability of finding X in an interval $[x_a, x_b]$ is given by $P(x_b) - P(x_a)$.

In case of having a discrete variable X_i with probability p_i ($i = 1, ..., N$, given N different states of X_i), the PDF can be described as

$$p(x) = \sum_{i=1}^{N} p_i \delta(x - i), \tag{1.35}$$

and, equivalently, the CDF is a piecewise constant function,

$$P(x) = \begin{cases} 0 & if\ x < 1 \\ \sum_{i=1}^{x} p_i & if\ 1 \leq x \leq N \\ 1 & if\ x > N \end{cases} . \tag{1.36}$$

1.5.2 Random sampling methods for probability density functions

In MC methods, we often need to sample different PDFs (e.g. a Gaussian distribution, the distribution of points on the suface of a sphere, or any arbitrary discrete distribution). Nonetheless, common random number generators are based on uniform distributions. For this reason, further methods have been developed to transform a random variable within a uniform distribution into other distributions [Kalos and Whitlock, 2008].

In this section, we review the basics of random number generators in subsection 1.5.2.1, and a few of the methods used to generate other PDFs in the following subsections 1.5.2.2, 1.5.2.3, 1.5.2.4, and 1.5.2.5.

1.5.2.1 Random number generators

Common random number generators are based on uniform distributions in the range $[0, 1)$. The generated sequences are actually called "pseudo-random", since they are based on deterministic algorithms that compute the next random number from the previous numbers in the sequence [James, 1990, L'Ecuyer, 1988, Marsaglia et al., 1990]. The algorithms implemented for this purpose are based on repetition cycles that are

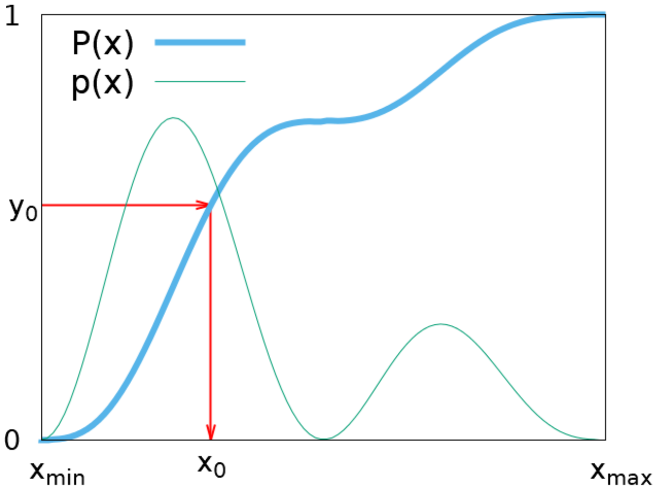


Figure 1.11: Graphical visualization of an example of the inverse transform method. The sampled PDF is given by $p(x)$, and its CDF is given by $P(x)$. A random sample of X, x_0, is obtained from a random sample of Y, $y_0 \in [0, 1)$, applying $x_0 = P^{-1}(y_0)$.

reasonably longer than the total numbers generated during the simulation [Agency, 2019]. Thus, the correlations in the sequence are not noticeable in the final MC results. The starting number of the sequence generated is based on one or a few seed numbers that define the whole series. We can change the seeds to generate different MC results of the same input distribution.

In the following subsections we present different random sampling methods based on the uniform distributions generated in the range $[0, 1)$.

1.5.2.2 Inverse transform method

The inverse transform method can be implemented to sample a given PDF of the variable X, $p(x)$, when the inverse of its CDF can be defined as the PDF of a variable Y, $p_y(y) = P^{-1}(y)$. The PDFs of Y and X are related by the expression $p(x)dx = p_y(y)dy$, thus having $p_y(y) = 1$ (uniform distribution in the range $(0, 1)$) [Agency, 2019, Kalos and Whitlock, 2008]. A random sample of $y_0 \in [0, 1)$ can be generated, and use $x_0 = P^{-1}(y_0)$ to find a sample of X with the distribution $p(x)$. In figure 1.11 we show a visual example of the method.

The drawback of the inverse transform method is that we may find functions with non analytical expression for $P^{-1}(y)$, thus limiting its performance to the numerical implementation. In [Agency, 2019], the rational inverse transform with aliasing (RITA) algorithm is presented with this purpose, using an optimal grid to interpolate the probability density function.

1.5.2.3 Rejection method

Rejection methods are a general technique to generate a sample of random numbers from a given distribution [Agency, 2019, Kahn, 1954, Kalos and Whitlock, 2008, Marsaglia et al., 1964]. These methods consist on a trial and test procedure that is repeated until a value is accepted. They have the advantage that there is no need to know the PDF normalization, and they adapt for any distribution without computations needed prior to the sample generation. In contrast, they might be very slow, since many values might be rejected before acceptance. A graphical example is shown in figure 1.12. First, we need to define a function $g(x) > p(x)$, from which we can sample a random value of x_0 (we can use the uniform function $g(x) = max(p(x))$ for general use, but it might not be optimal). A second random number is generated uniformly in the range $[0, g(x_0)]$, and we accept x_0 with probability $p(x_0)/g(x_0)$. The process is repeated until a value is accepted (for the example in fig. 1.12, the value is rejected). The method is summarized here:

1. Generate a random number x_0 from the distribution $g(x)$.
2. Generate $\xi \in [0, 1)$, set $y_0 = \xi \cdot g(x_0)$.
3. $\begin{cases} Accept\ x_0 & if\ y_0 < p(x_0) \\ Back\ to\ step\ 1 & if\ y_0 > p(x_0) \end{cases}.$

The probability to obtain an acceptance, that is, the algorithm efficiency, from the ratio between the areas under $p(x)$ and $g(x)$ can be computed as

$$\epsilon = \frac{\int_{x_{min}}^{x_{max}} p(x)dx}{\int_{x_{min}}^{x_{max}} g(x)dx}. \tag{1.37}$$

The expected number of trials up to an acceptance is given by $1/\epsilon$.

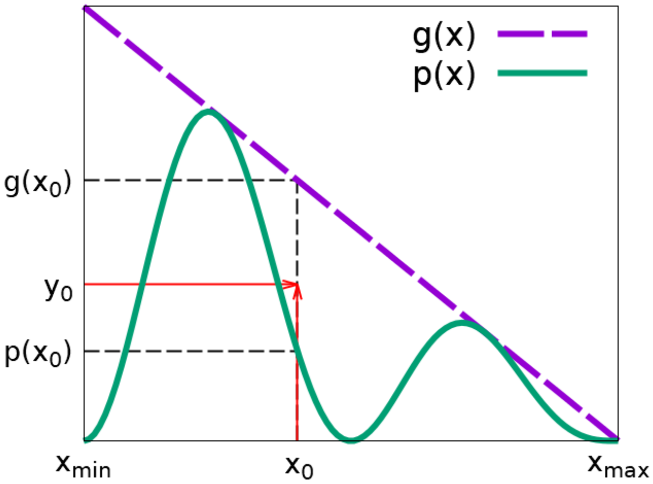


Figure 1.12: Graphical visualization of an example of the rejection method. The PDF to be sampled is $p(x)$, and $g(x)$ is defined so that it satisfies $g(x) > p(x)$. We first sample x_0 from $g(x)$, and y_0 is sampled uniformly in the range $[0, g(x_0)]$. In this case, the value is rejected since $y_0 > p(x_0)$.

1.5.2.4 Sampling discrete distributions

The inverse transform method in 1.5.2.2 and the rejection method in 1.5.2.3 have been defined for continuous distributions, but very often we may need to sample discrete PDFs. In a general framework, both might be adapted to satisfy a discrete distribution as it is defined in equation 1.35.

In the case of inverse transform methods, a simple way to invert a discrete CDF is to iterate over $i = 1, ..., N$ until we find that the condition $P_i = \sum_{i'=1}^{i} p_{i'} > \xi$ is satisfied (given that ξ is uniformly distribute in $[0, 1)$). The output x will correspond to the random sample. In [Agency, 2019], the authors proposed to use a binary search instead of a sequential search for distributions with a large number of possibilities N, thus effectively reducing the number of comparisons needed to find x. Another practical solution would be the use of a lookup table that relates a random sample of ξ to a reduced set of values of $x \in [i_0, i_f]$. Maximum efficiency is obtained for a one to one coupling with x.

For the case of rejection methods, it is sufficient to define a uniform distribution $g(x) = max(p_i)$, and define x as the integer part of $\xi N + 1$. But the algorithm might become very inefficient for certain distributions.

In the following section we describe an alternative method with a single comparison needed to obtain each random sample.

1.5.2.5 Walker's aliasing method

The Walker's aliasing method [Walker, 1977] is an efficient algorithm for discrete distributions sampling. In this section we explain the approach from [Agency, 2019, Salvat, 1987], who reformulated the algorithm using a single random number (two were used originally). In figure 1.13, a graphical representation of this procedure is shown. Suppose we want to pick a random pixel $i \in [1, N]$ with normalized probability density function p_i. We can distribute the function p_i in a bars histogram with height Np_i and width $1/N$. From this histogram, we rearrange the bars distribution to make a square of unit area, cutting the biggest bar to complete the smallest one until they all have height 1. In this process we keep the "alias" value K_l of the bar position that was used to complete each single l-th position, and the cutoff value F_l of the original l-th bar. It is clear that we can define a random point from the distribution p_i if we sample the unit square with two random variables. Instead, we add the bars one after the other to obtain a segment of length N, from which we can select a random point generating a single random number ξN (where $\xi \in [0, 1)$):

1. Generate $\xi \in [0, 1)$, and set $R = \xi N + 1$.
2. Set $j = int(R)$ and $r = R - j$.
3. $i = \begin{cases} j & if\ r < F_i \\ K_j & if\ r \geqslant F_i \end{cases}$.

As we can see, the cutoff values F_j represent the probability to pick the voxel j, whereas the alias K_j is chosen with probability $1 - K_j$. The total probability p_i of each point may be reconstructed from the alias and cutoff values,

$$Np_i = F_i \sum_{j \neq i} (1 - F_j)\, \delta_{i,K_j}, \tag{1.38}$$

where δ_{i,K_j} is the Kronecker delta.

We see that the algorithm gets a sample of p_i with a single call to the random numbers generator and a single comparison, at the expense of double memory usage

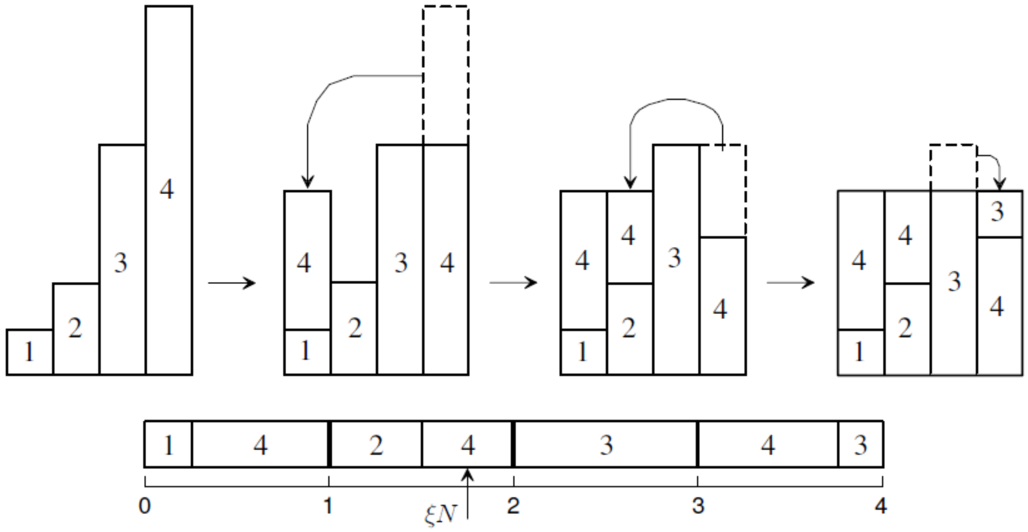


Figure 1.13: Graphical visualization of the Walker's method [Agency, 2019]. The height of the bars represent the discrete probability of obtaining a bin from 1 to 4. Once they are reorganized in a square of unit area, we can extend it in a vector to sample the original distribution.

(two vectors, one for the aliases and one for the cutoff values). The inverse transform method implies a high number of comparisons, even if the search is optimized, and other techniques, like rejection methods [Agency, 2019, Kalos and Whitlock, 2008], may require a high number of comparisons and random number generations before accepting a sample.

1.5.3 Variance reduction methods

In general, MC simulations might be computationally expensive, and different methods might be approached to increase their efficiency [Agency, 2019, Kalos and Whitlock, 2008]. Basic variance reduction techniques consist on avoiding unnecessary calculations. We can avoid the sampling of a certain space of the PDF that we know is not of interest for the final result, for example, restricting the emission solid angle of a radioactive source only to the area where the detector is located, and set a weight for each particle equivalent to the relative solid angle coverage. If this is the case, we must be sure that the particles will travel in straight lines. Other simple method is to avoid simulating non-useful particles. In chapter 5, secondary particles coming from photon interactions are not considered at all, and positron simulation is substituted for a convolution method.

More sophisticated solutions may include interaction forcing, splitting particles, or Russian roulette techniques [Agency, 2019]. These methods involve increasing the chances of a given particle to achieve low probability events, using weights to maintain the quantities to be computed unbiased. In contrast, this technique will increase the variance in other quantities.

1.6 High-performance computing with GPUs

In 1965, Gordon E. Moore experimentally postulated that the number of processors in a transistor would exponentially raise with time, doubling every 18 months [Moore, 2006]. Application developers have benefited from Moore's law to improve complex algorithms until very recently, when CPUs manufacturers seem incapable to maintain the trend any more [Pratx and Xing, 2011]. In contrast, the computer graphics industry demand of high parallel computing tasks has boosted the development of graphics processing units (GPUs) with thousand of processors since the 1990s [Fatica and Ruetsch, 2014]. Thanks to the appearance of application program interfaces (APIs) such as NVIDIA CUDA, OpenCL, or OpenACC, that provide a user-friendly environment, GPU programming has been tremendously adopted in both industry and academy.

In medical physics, GPU programming has gained noticeable interest in the recent years. As stated in Després and Jia [2017], the acceleration obtained translates to higher speed in clinical practice workflows, higher accuracy in the results achieved thanks to more computationally complex algorithms, and management of higher loads of data to solve large size problems in acceptable times. In consequence, it has become a standard that any competitive software needs to be accelerated with massive parallel computing hardware, among which GPUs are mostly used. A wide variety of accelerated models for image reconstruction [Després and Jia, 2017, Pratx et al., 2009], image processing [Eklund et al., 2013], and dose planning [Pratx and Xing, 2011] have emerged. Many authors have also adapted Monte Carlo codes to GPU parallel computing [Alerstam et al., 2008b, Badal and Badano, 2009, Bert et al., 2013, Chi et al., 2016, Galve et al., 2020a, Hissoiny et al., 2011, Ibáñez et al., 2021, Jia et al., 2011, Lai et al., 2019, Lippuner and Elbakri, 2011, Ma et al., 2020]. In section 1.4.4 we have already discussed about the progress of machine learning algorithms thanks to parallel computing.

The GPU programming is the workhorse of this thesis. All the applications developed rely on the foreground of parallel computing acceleration with GPUs and CUDA Fortran [Fatica and Ruetsch, 2014, NVIDIA, 2021], a compiler jointly developed by NVIDIA and the Portland Group to get Fortran codes running on the GPU (although the most widely extended is CUDA C [Sanders and Kandrot, 2010], in practice both operate at a similar level of parallel optimization). In chapter 3 we give the details of a flexible reconstruction software, GPU-PET, accelerated with GPUs, and chapter 5 details the development and performance of a GPU accelerated PET simulator. Chapters 4 and 6 rely on these tools to improve PET image quality.

1.6.1 CUDA GPU architecture and programming

When programming a GPU there are a few considerations to account for that differ from CPU programming. We are going to focus in the CUDA architecture from NVIDIA [Fatica and Ruetsch, 2014, Sanders and Kandrot, 2010], since it is the one used in this thesis. CUDA is a hybrid programming model of CPU and GPU execution, in which the GPU kernels are called from the CPU. Usually we refer to the CPU as the host, and the GPU as the device. In the device kernels, we can launch thousands of threads to perform the same routine in parallel, but operating over different memory locations. Optimal performance is achieved when adjacent threads work on adjacent chunks of data (e.g. contiguous elements of an array), concept known as coalescence. The threads are grouped into blocks, and the blocks are grouped into grids, so that every thread can be identified by its thread index and its block index (see figure 1.14). Inside a block, the threads are grouped in warps of 32 threads that work in single-instruction multiple thread (SIMT) fashion. This means that threads within a warp perform the instructions in a lockstep. For example, in the case of *if-else* statements, the threads within a warp will execute sequentially all the conditional blocks, thus enlarging the time consumption. It is also important to mention the case of multiple threads writing on the same memory location at a time, leading to non deterministic outputs. Atomic operations ensure that no more than one thread can update a memory location at a time at the expense of additional computational time.

The communication from the GPU threads to the CPU is done through different memory spaces [Fatica and Ruetsch, 2014, NVIDIA, 2021]. In figure 1.14 we graphically show the access to these spaces. First we have to differentiate between the dynamic

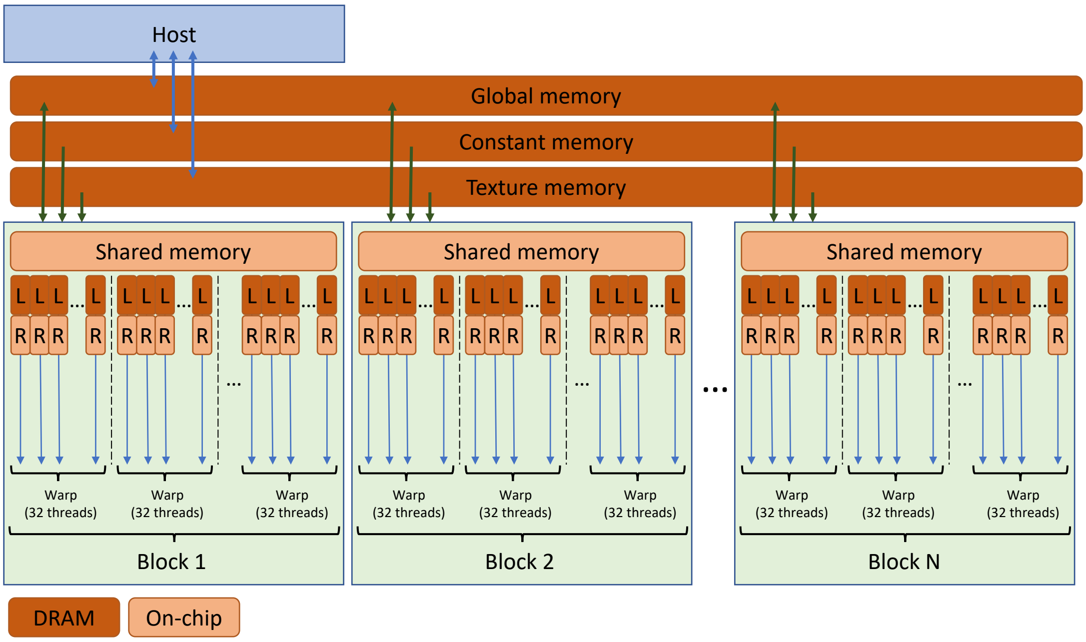


Figure 1.14: Graphical visualization of the GPU CUDA architecture and memory access from every thread. The physical location of each memory space in the device is shadowed in dark orange (dynamic random-access memory) and light orange (on-chip). The host has input/output (I/O) access to the global, constant, and texture memory. All the blocks have input access to these spaces, but they can only write in the global memory. Inside each block we find shared memory, local memory (L), and the registers (R). The shared memory is I/O accessed by all the threads within a block. The local memory and the registers are limited I/O memory spaces to every thread. The threads are internally grouped in warps of 32 threads within every block.

random-access memory (DRAM) and on-chip memory. DRAM (dark orange in figure 1.14) can be accessed from the host and from the device, but it is off-chip in the sense that it is physically separated from the GPU thread processors. In DRAM we find global memory, constant memory, texture memory and local memory:

- The **global memory** can be read and written by the host and by any thread, and it is the communication bridge between host and device.
- **Constant memory** and **texture memory** are read-only spaces from the device, and they can only be written by the host. Both are cached on the GPU for fast access, but use different pathways for accessing global memory. The constant memory space is very limited.

- **Local memory** is thread private memory limited to the thread lifetime, that can be either allocated on the global memory, or on the registers (on-chip memory).

The on-chip memory is located on the GPU multiprocessors (groups of 32 thread processors), and it can only be accessed by the thread processors in each given multiprocessor. Here we find the shared memory and the registers:

- The **shared memory** is block private memory, only accessible for the threads within a block. It can be used for thread-communication within a block.
- The **registers** are assigned per-thread, and they are mainly dedicated to accelerate access to local memory (optimal performance is achieved when local memory is reduced to scalars saved in the registers).

As a rule of thumb, accessing constant memory, texture memory, shared memory, and registers is faster than accessing global memory, but global memory is the largest space of memory and the only output space to bring results back to the host.

Chapter 2

Materials and tools

In this chapter, the hardware and software tools used in this thesis are summarized. In section 2.1 we will give the details of the simulation code PeneloPET, section 2.2 is dedicated to the scanners employed for simulations and real data acquisitions in the thesis, and sections 2.4, 2.3 and 2.5 are dedicated to the real and simulated phantoms employed.

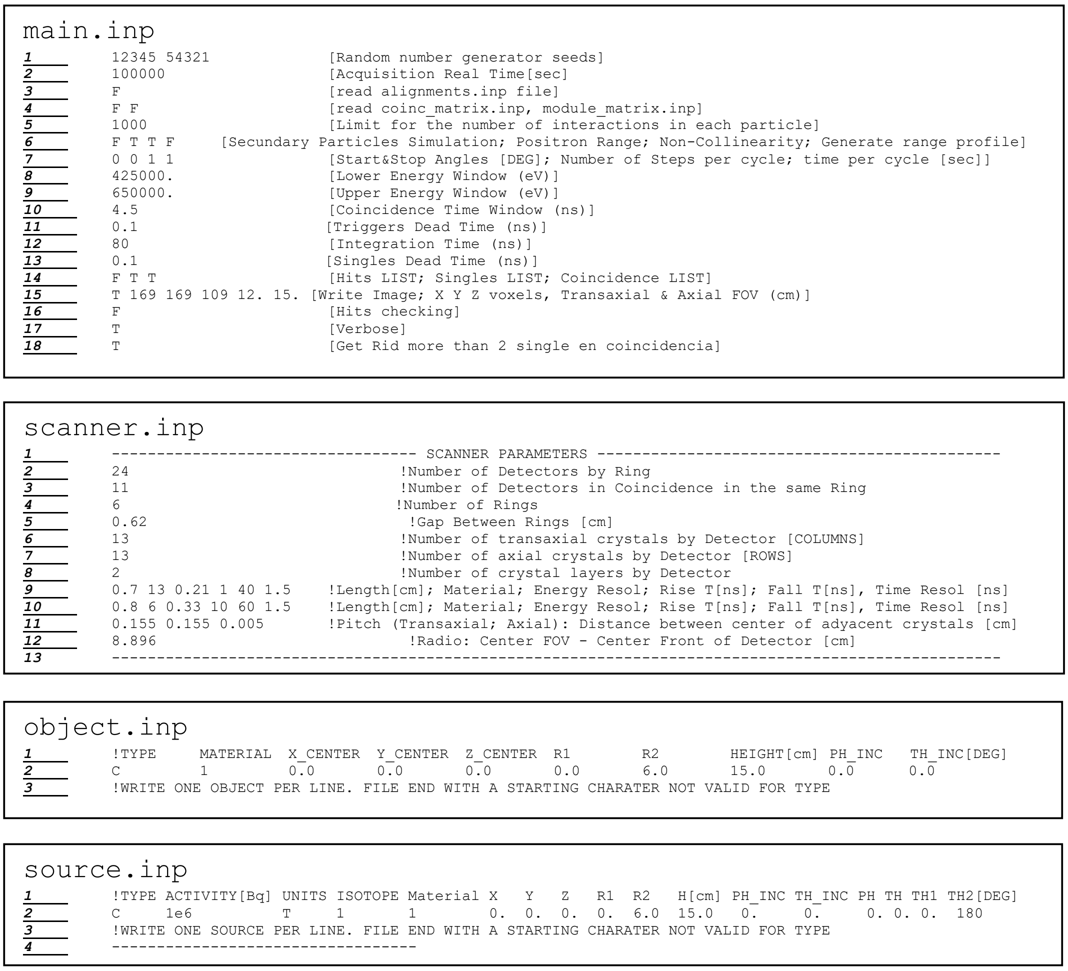
2.1 PeneloPET

Among the wide variety of Monte Carlo (MC) PET dedicated simulators available, we have used PeneloPET[1]as a standard reference [Lopez-Montes et al., 2019] (see section 5.1 for further discussion about MC simulators in PET), which has been developed in the Nuclear Physics Group at University Complutense of Madrid. PeneloPET has been extensively validated in different studies against real data [Abushab et al., 2016, Cal-González et al., 2013, 2015a, España et al., 2009, Popota et al., 2015].

PeneloPET has been built on the basis of PENELOPE, a MC tool implemented in the Fortran language for the transport of electrons, positrons and photons in the energy range of 1 keV to 1 GeV. In the PeneloPET package, we find a user-friendly environment based on simplified input text files to define the simulation parameters (time duration, inputs and outputs, etc.), the scanner geometry (detector module and scanner geometry), the emission source, and the objects in the medium. In figure 2.1, an example of the input files needed to define a simulation of the 6R-SuperArgus scanner

---

[1] http://nuclear.fis.ucm.es/penelopet/



main.inp

```
12345 54321             [Random number generator seeds]
100000                  [Acquisition Real Time[sec]
F                       [read alignments.inp file]
F F                     [read coinc_matrix.inp, module_matrix.inp]
1000                    [Limit for the number of interactions in each particle]
F T T F     [Secundary Particles Simulation; Positron Range; Non-Collinearity; Generate range profile]
0 0 1 1                 [Start&Stop Angles [DEG]; Number of Steps per cycle; time per cycle [sec]]
425000.                 [Lower Energy Window (eV)]
650000.                 [Upper Energy Window (eV)]
4.5                     [Coincidence Time Window (ns)]
0.1                     [Triggers Dead Time (ns)]
80                      [Integration Time (ns)]
0.1                     [Singles Dead Time (ns)]
F T T                   [Hits LIST; Singles LIST; Coincidence LIST]
T 169 169 109 12. 15. [Write Image; X Y Z voxels, Transaxial & Axial FOV (cm)]
F                       [Hits checking]
T                       [Verbose]
T                       [Get Rid more than 2 single en coincidencia]
```

scanner.inp

```
---------------------------------- SCANNER PARAMETERS ---------------------------------------------
24                              !Number of Detectors by Ring
11                              !Number of Detectors in Coincidence in the same Ring
6                              !Number of Rings
0.62                             !Gap Between Rings [cm]
13                              !Number of transaxial crystals by Detector [COLUMNS]
13                              !Number of axial crystals by Detector [ROWS]
2                               !Number of crystal layers by Detector
0.7 13 0.21 1 40 1.5    !Length[cm]; Material; Energy Resol; Rise T[ns]; Fall T[ns], Time Resol [ns]
0.8 6 0.33 10 60 1.5    !Length[cm]; Material; Energy Resol; Rise T[ns]; Fall T[ns], Time Resol [ns]
0.155 0.155 0.005       !Pitch (Transaxial; Axial): Distance between center of adyacent crystals [cm]
8.896                            !Radio: Center FOV - Center Front of Detector [cm]
----------------------------------------------------------------------------------------------------
```

object.inp

```
!TYPE     MATERIAL  X_CENTER  Y_CENTER  Z_CENTER  R1        R2        HEIGHT[cm] PH_INC   TH_INC[DEG]
C         1         0.0       0.0       0.0       0.0       6.0       15.0       0.0      0.0
!WRITE ONE OBJECT PER LINE. FILE END WITH A STARTING CHARATER NOT VALID FOR TYPE
```

source.inp

```
!TYPE ACTIVITY[Bq] UNITS ISOTOPE Material X   Y   Z   R1  R2   H[cm] PH_INC TH_INC PH TH TH1 TH2[DEG]
C     1e6          T     1       1        0.  0.  0.  0.  6.0  15.0  0.     0.     0. 0. 0.  180
!WRITE ONE SOURCE PER LINE. FILE END WITH A STARTING CHARATER NOT VALID FOR TYPE
----------------------------------
```

Figure 2.1: Example of the input files required to simulate the 6R-SuperArgus scanner (see section 2.2.1) with a uniform cylindrical source of 6 cm radius and 15 cm length inside.

from section 2.2.1 is shown. By default, the simulator defines cylindrical scanners, although it allows for simple modifications in this geometry. Detector response can be modeled in detail, including the position, energy and timing response of the detector and electronics. The positron range profiles and decay schemes of many β^+ radionuclides are available (^{18}F, ^{11}C, ^{13}N, ^{15}O, ^{82}Rb, ^{124}I, and ^{68}Ga), and others can be easily generated by the user. Different output levels are available, from particle interaction or single events list-mode files, to coincidences list-mode files or sinograms.

During the development of this thesis, the author collaborated in the development of the most recent version of PeneloPET [Lopez-Montes et al., 2019], and many tests have been performed with this platform to assess scanner performance and different reconstruction methods. In this manuscript, we used PeneloPET to validate the UMC-PET simulator of chapter 5.

2.2 PET scanners

2.2.1 The SuperArgus family of scanners

In this section we are going to give the details of the SuperArgus family of scanners (Sedecal Medical Imaging). These scanners are based on the SuperArgus detector module (based on the GE Healthcare eXplore Vista from General Electrics [Wang et al., 2006]), which can be assembled in several configurations. The SuperArgus module is conformed of detector units of pixellated arrays of 13×13 crystals of 1.55 mm crystal pitch (0.1 mm reflector thickness), with a dual layer phosphor sandwich (phoswich) strategy for depth of interaction (DOI) information. The phoswich couples are made of lutetium-yttrium orthosilicate (LYSO) crystals of 7 mm length in the front layer optically joined to cerium-doped gadolinium orthosilicate (GSO) crystals of 8 mm length in the rear layer. The difference in light decay time (LYSO, 40 ns, GSO, 60 ns) is used to distinguish the interaction point in each layer, and the light output of the phoswich ensemble is coupled to position-sensitive photomultiplier tubes (PMT) to identify the crystal in the 13×13 arrays. The energy resolution is 21% in the LYSO crystals, and 33% in the GSO crystals (aggregated energy resolution is 24%), and the time resolution is 1.5 ns. The energy window used for the acquisitions in this thesis is 425-600 keV, and the coincidence time window is of 4.5 ns.

In this thesis, we have used the scanners with 6 and 4 rings of 24 detectors, so called 6R-SuperArgus and 4R-SuperArgus, and the 2R-Argus scanner with 2 rings of 18 detectors, commercialized for several years as the GE Healthcare eXplore Vista (General Electrics) [Wang et al., 2006]. For the 6R- and 4R-SuperArgus, the ring diameter is 17 cm, and the commercial transverse FOV is 12 cm diameter. Each ring of detectors need to leave an axial gap of 6.4 mm (∼4 crystals) because of the PMT width, thus having 15 cm and 10 cm axial FOV respectively. In the case of the 2R-Argus scanner, the ring diameter is 11.8 cm, the transverse FOV is 7 cm and the axial FOV is 4 cm. We will work without a maximum ring difference limit. In figure 2.2 a schematic representation of the three scanners is shown.

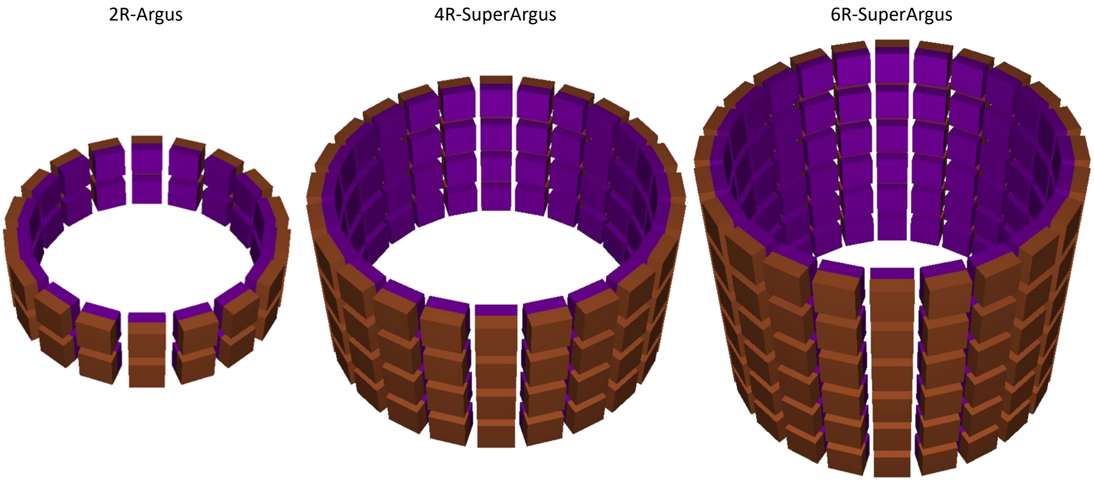


Figure 2.2: Scheme of the LYSO-GSO phoswich detectors (purple and brown, respectively) in the 2R-Argus (left), 4R-SuperArgus (center), and 6R-SuperArgus scanner (right). Images obtained from PeneloPET.

The results from the 6R-Superargus are shown in chapters 3 and 6, and the scanner was simulated in the section 5.3 of chapter 5. The 4R-SuperArgus scanner was used in chapter 4. The 2R-Argus scanner is simulated in chapter 6.

2.2.2 The Biograph mMR scanner

The Biograph mMR scanner is a PET/MRI compatible scanner for whole-body PET imaging [Delso et al., 2011]. The detector units are based on pixellated arrays of 8×8 lutetitum oxyorthosilicate (LSO) crystals of 4 mm crystal pitch and 20 mm depth. The light readout is performed by an array of 3×3 avalanche photodiodes (APD). The energy resolution is 14.5% and the energy window is 430-610 keV. The time resolution is 2.93 ns, and the coincidence window is of 5.86 ns. The detectors assembly consists of 8 rings on 56 detectors with 65.6 cm diameter, resulting in 59.4 cm transverse FOV and 25.8 cm axial FOV. The scanner can acquire static multibed data in 3D mode with delayed window random estimation. In figure 2.3, a schematic representation of the scanner is shown.

This scanner utilizes a sinogram format with maximum ring difference of 60 crystals and span 11. A real sinogram is shown in figure 2.4. The sinogram has 11 segments with 837 slices (127 direct slices in segment 0, and subsequently 115 in ±1, 93 in ±2,

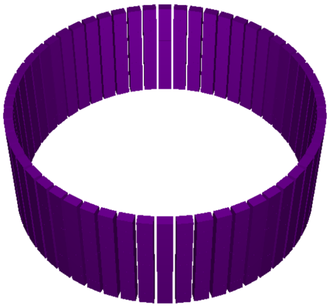


Figure 2.3: Scheme of the LSO detectors (purple) in the Biograph mMR scanner. A cylindrical object of 60 cm diameter is placed inside the scanner bore (brown object). Image obtained from PeneloPET.

71 in ±3, 49 in ±4, and 27 in ±5). The radial and angular bins do not follow a linear relationship with physical distances and angles, but a direct correspondence with every couple of crystals in the plane. The format has been specified by the manufacturer.

We worked with data acquired in the Biograph mMR in section 5.4 of chapter 5. Results from simulations of this scanner are shown in chapter 3.

2.2.3 The High Spatio Temporal Resolution BrainPET prototype scanner

The High Spatio Temporal Resolution BrainPET (HSTR-BrainPET) is a prototype PET scanner for brain dedicated studies [Catana, 2019a,b, Galve et al., 2020b]. The prototype has a spherical morphology to enlarge its sensitivity, and it is designed to be integrated with a 7-Tesla MRI scanner. We show a scheme of the prototype in figure 2.5. In this thesis, the scanner was simulated since it has not been manufactured yet. The simulated spherical geometry has 32 cm inner diameter, with 880 detector blocks facing the center of the sphere. The scanner has two open sides, one for the insertion of the patient's head, and other for a PET compatible MRI coil insert. We have defined pixelated detectors of LSO of 8×8 crystal units with 2×2 mm^2 cross section, and 26 mm length. Since the multiple layer structure of the detectors was not yet defined at the time of this study, we have compared the results with detector blocks of 9 mm length to assess the DOI effect. The energy resolution for 511 keV is 12%, and the energy window implemented is 400-700 keV.

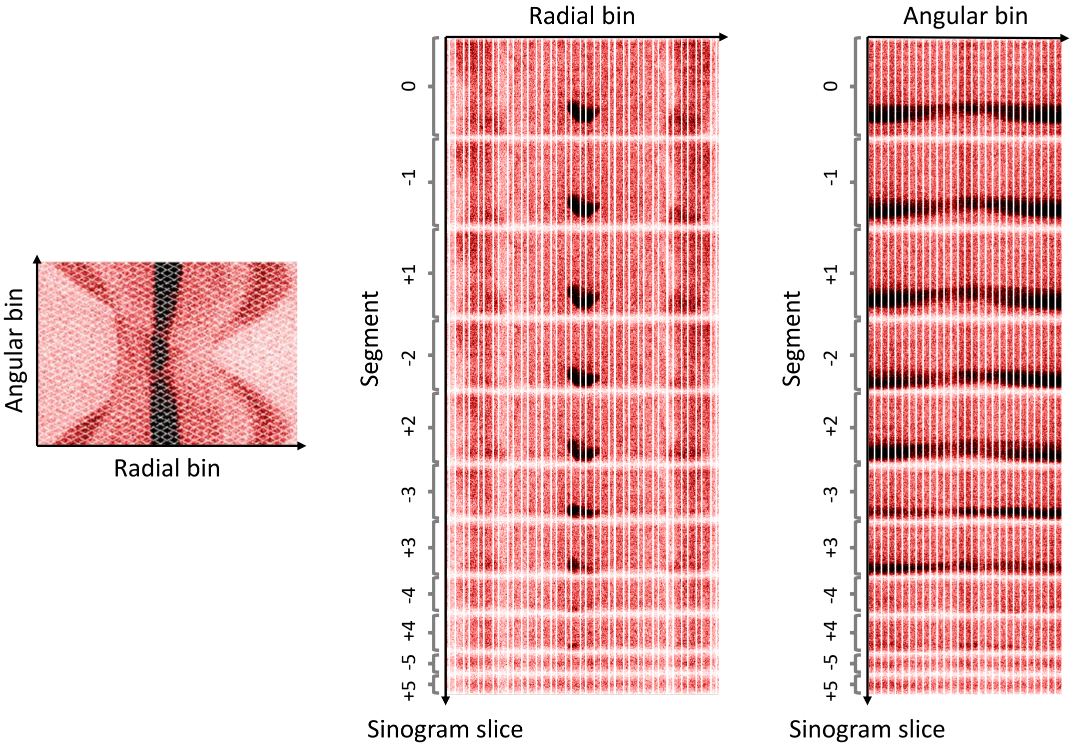


Figure 2.4: Example of one of the sinograms from the Biograph mMR utilized in section 5.4 of chapter 5.

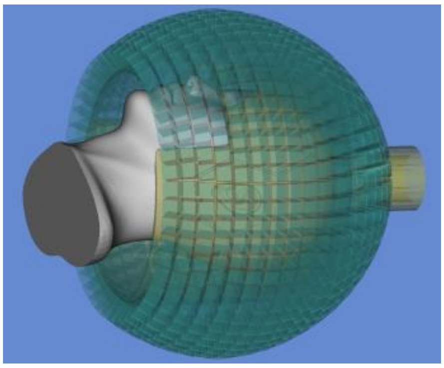


Figure 2.5: Graphical representation of the HSTR-BrainPET prototype scanner.

2.3 Image Quality phantom and evaluation of the NEMA NU4-2008 protocol

The NEMA NU4-2008 protocol National Electrical Manufacturers Association [2008] establishes standard procedures to assess the system performance for PET preclinical scanners. This section corresponds to the section 6 of the original document. Here we summarize the details about the image quality (IQ) phantom, activity and acquisition (section 2.3.1), and the methods to analyze the image (section 2.3.2) through noise or percent standard deviation (%STD), recovery coefficients (RC), and spill over ratio (SOR).

2.3.1 The Image Quality phantom

Following the protocol, the activity employed for the total phantom should be of 100 µCi within ±5% of ^{18}F acquired for 20 minutes. The phantom is made of PMMA (polymethyl methacrylate) and it has a size of 30 mm inner diameter and 50 mm length. A schematic view of the different sections of the phantom is presented in 2.6. The phantom is composed of a body (subfigures 1 and 4 in 2.6), an upper cover (subfigures 2 and 5 in 2.6), and a bottom cover (subfigures 3 and 6 in 2.6). The body phantom is the only region that is filled with uniform activity.

Once the three parts are attached and filled, the phantom offers three main regions of study. First, a uniform region in the middle of the phantom, with 30 mm diameter and 15 mm length (between the downside small rods region and the upper cylinders of the top cover). In this region we will measure the image noise. In the upper 15 mm of the phantom, we find the two cylinders of the top cover. One of the cylinders will be empty, and the other is filled with water (without activity). Here we measure the SOR, that estimates the quality of attenuation and scatter correction. Finally, in the down side of the phantom we find five active rods of diameters from 1 mm to 5 mm that are used to measure the image RC.

2.3.2 The image analysis

The image quality is evaluated with three parameters: percent standard deviation (%STD), recovery coefficients (RC), and spill over ratio (SOR).

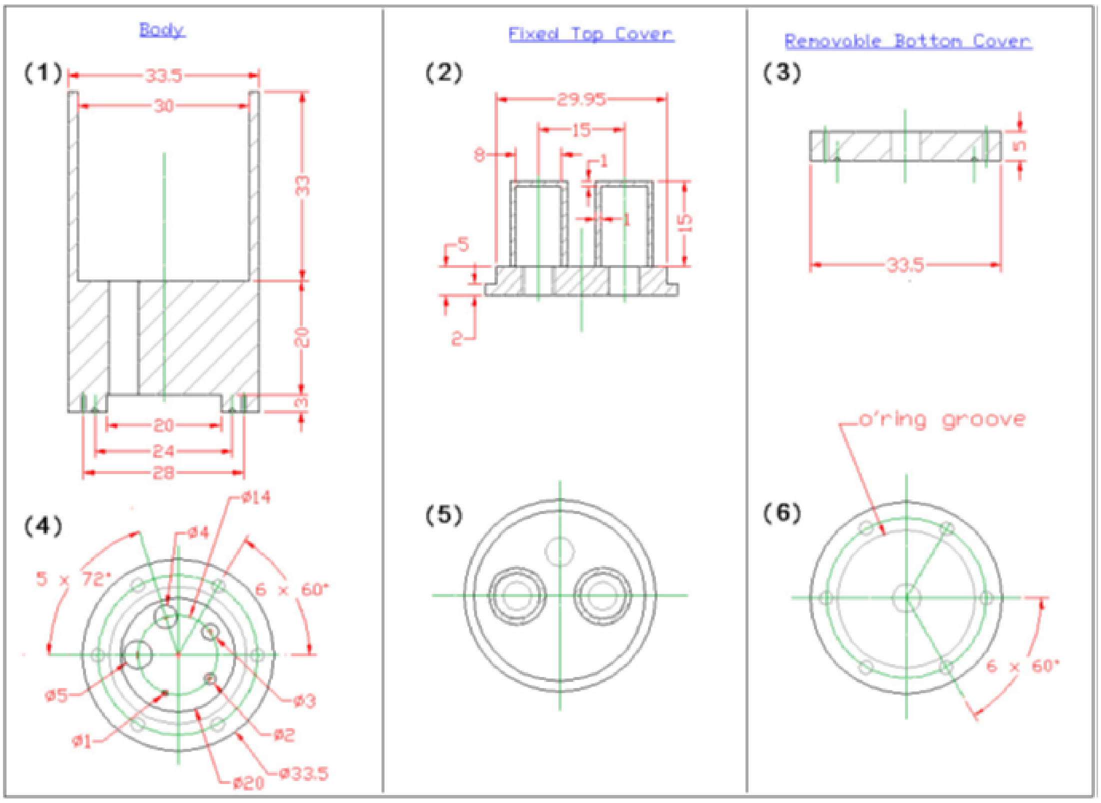


Figure 2.6: Coronal view of the image quality phantom body (1), the top cover (2) and the bottom cover (3), and transverse views respectively (4), (5) and (6). All dimensions are in mm. Image taken from the NEMA NU4-2008 protocol National Electrical Manufacturers Association [2008].

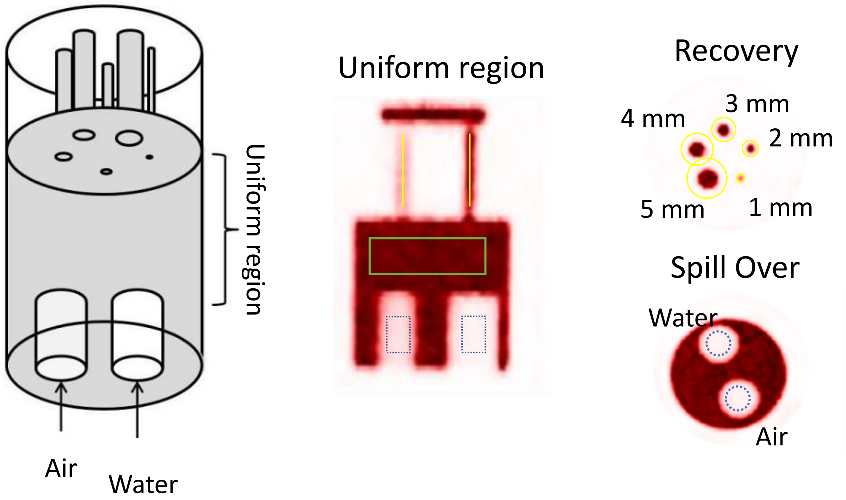


Figure 2.7: Scheme of the image quality phantom of the NEMA NU 4-2008 National Electrical Manufacturers Association [2008] (left) and regions for noise (center, green region in the uniform region of the phantom), recovery coefficients (right top, yellow regions in the rods; transverse top view), and spill over (right bottom, blue regions in the non-active inserts filled with water and air; transverse bottom view) quantification.

2.3.2.1 Percentage Standard Deviation

Centered within the uniform region of the phantom, we draw a cylindrical region of interest (ROI) of 22.5 mm diameter and 10 mm long (see figure 2.7). The percentage standard deviation (%STD, also called noise) is measured as the ratio between the standard deviation and the average activity inside the background region,

$$\%STD = 100 \times \frac{STD_{background}}{Mean_{background}}. \tag{2.1}$$

2.3.2.2 Recovery coefficient

First, we collapse the 10 mm central slices in the small rods, and for each rod we search for the maximum pixel within a circular region of twice the rod diameter. We draw an axial line profile at the transverse position of the maximum pixel for the 10 mm slices (see figure 2.7). The ratio between the average activity in the profile and the average activity in the background region (from previous section 2.3.2.1) is the RC,

$$RC = \frac{Mean_{profile}}{Mean_{background}}. \tag{2.2}$$

2.3.2.3 Spill Over Ratio

We draw two cylindrical ROIs of 4 mm diameter and 7.5 mm length in the water and air cylinders of the top cover (see figure 2.7). The ratios of mean activity in each region and the background region is defined as the SOR,

$$\begin{aligned} SOR_{water} &= \frac{Mean_{water}}{Mean_{background}} \\ SOR_{air} &= \frac{Mean_{air}}{Mean_{background}}. \end{aligned} \tag{2.3}$$

2.4 Derenzo phantom

Derenzo phantoms [Derenzo et al., 1977] are commonly used in PET to assess image resolution, and some authors have proposed to include them in the image quality standards [Hallen et al., 2020]. A Derenzo phantom is composed of parallel rods in triangle distribution, filling different sections of a cylinder, each with a different diameter. A real Derenzo phantom with 45 mm diameter and rods diameters ranging from 1.2 mm to 4.8 mm was employed in chapters 3 and 4. This phantom has been acquired both with hot and cold rods. Further simulated Derenzo phantoms are described in section 2.5.1.

2.5 Simulated phantoms

In this section we summarize the voxelized phantoms that were simulated along this thesis. Among them we find Derenzo phantoms (2.5.1, a Cologne phantom 2.5.2, and a Zubal phantom 2.5.3. All the simulations were ran on the UMC-PET simulator of chapter 5.

2.5.1 Derenzo phantoms

In this work we have simulated two Derenzo phantoms, both with hot rods over a mild background. In chapter 3, the rods to background ratio is 10:1, with diameters ranging

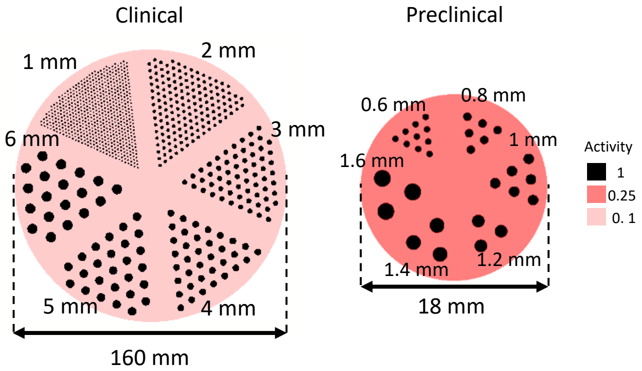


Figure 2.8: Activity distribution of a clinical (left) and a preclinical (right) Derenzo phantoms simulated in this thesis. Notice that the size scale are not proportional. In the clinical phantom, the rods to background ratio is 10:1, whereas in the preclinical one it is 4:1.

from 1 mm to 6 mm diameter, total phantom size is 160 mm diameter and 50 mm length (left-hand side of figure 2.8). In chapter 6, a smaller Derenzo phantom with background activity of one fourth the activity in the rods, 18 mm diameter, 10 mm depth, and 6 sections of rods of diameters 1.6 mm, 1.4 mm, 1.2 mm, 1 mm, 0.8 mm and 0.6 mm was simulated (right-hand side of figure 2.8).

2.5.2 The Cologne phantom

Similar to the Derenzo phantom, the Cologne phantom [Kotasidis et al., 2011] can be used to assess the scanner resolution. The Cologne phantom consists on 5×5 squares of point sources of different diameters, ranging from 2 mm to 5 mm (see figure 2.9). This phantom was used in chapter 3.

2.5.3 The Zubal phantom

The Zubal phantom used in this work is shown in figure 2.10. This phantom is voxelized in 87 organs that are categorized in seven tissues. In table 2.1 we summarize the activity distribution implemented for each tissue. The phantom has 256×256×128 voxels of 1.1×1.1×1.4 mm^3.

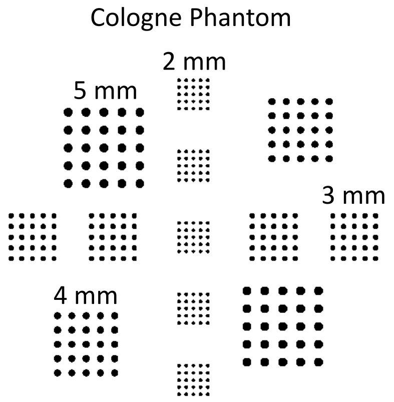


Figure 2.9: Activity distribution of the Cologne phantom.

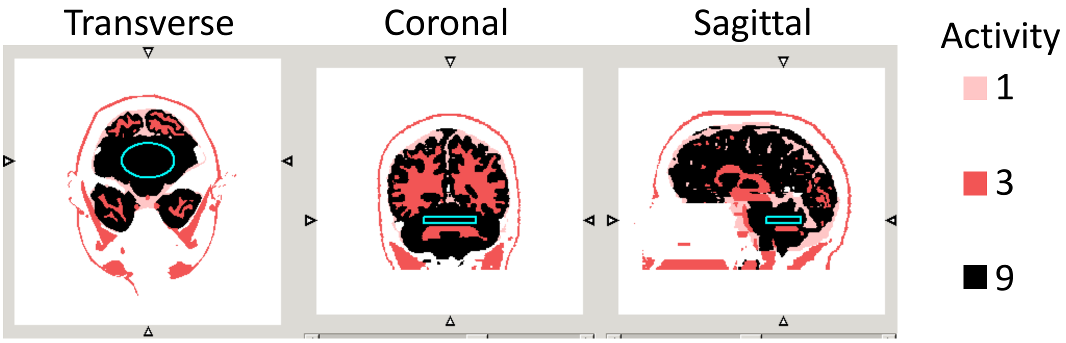


Figure 2.10: Activity distribution of the Zubal phantom. In the right-hand side we show the legend of the relative activity inside the phantom. Inside an homogeneous region of the cerebellum, we have drawn the elliptic cylinder marked in blue to measure noise in the reconstructed images.

Tissue	Activity
Soft tissue	3
Fat/adipose tissue	0
Cortical bone	0
Gray matter	9
White matter	3
Cerebrospinal fuid	1
Cartilage	0

Table 2.1: Relation of the activity assigned to each tissue in the Zubal phantom.

Chapter 3

GPU-based PET image reconstruction software

3.1 Introduction and motivation

Current scanners have increased the reconstruction complexity in terms of data size (because of axially extended field of view (FOV) [Nadig et al., 2021, Slart et al., 2021, Udias et al., 2018] and novel multi-layer detectors [Kang et al., 2021, Mohammadi et al., 2017]) and novel non-cylindrical geometry management [Catana, 2019b, Perez-Benito et al., 2018, Samanta et al., 2021, Tao et al., 2018, Yoshida et al., 2020] to improve sensitivity and optimize spatial resolution. One of the challenges of these larger scanners is the large computational cost of the projection and backprojection operations needed in the iterative reconstruction algorithms (see section 1.4.2). The advent of GPU parallel computing boosted the reconstruction process, enhancing acceleration of these two basic components of the MLEM algortihm [Barker et al., 2009, Herraiz et al., 2011, Markiewicz et al., 2014, Meng et al., 2019, Pratx et al., 2009, Zhou and Qi, 2011, 2014]. The System Response Matrix (SRM) modeling also requires larger memory sizes, and many authors employed symmetries [Cabello et al., 2010, Gong et al., 2017, Herraiz et al., 2006a, Scheins et al., 2011, Wei and Vaska, 2020], SRM sparseness principles [Li et al., 2015, Xu et al., 2019], and SRM decomposition [Pilleri et al., 2019, Zhou and Qi, 2014] to reduce it.

The purpose of this chapter is to explain the details of the GPU-based PET image reconstruction software (GPU-PET reconstructor) developed during this thesis. The code was designed to match optimal trade off between speed and flexibility, avoiding image quality detriment. One of the main goals was to develop a framework to define the scanner geometry without the constraints of traditional sinograms, limited to cylindrical shaped scanners. To face this challenge, we implemented a novel input format based on the coordinates of the crystal units conforming the scanner, and their relation with every LOR index. Speed is achieved thanks to the GPU implementation of the main MAP-OSEM [Alenius and Ruotsalainen, 1997, Hudson and Larkin, 1994, Nuyts et al., 2002] routines. The projection and backprojection routines are implemented in a LOR-driven (or ray-driven) approach, and the image filter routines use a voxel-driven scheme. These routines are designed in separate modules to allow for ease of modifications of the reconstruction scheme, the input data format (LOR histogram, sinogram, or list-mode data), or scanner geometry. The SRM response is decomposed in a projection model and a Gaussian image blurring kernel [Reader et al., 2003] to minimize memory storage. With this structure, we expect to face current reconstruction challenges, either for clinical, preclinical, or non-conventional scanners (see section 2.2) in a simple and flexible tool. Initial results were presented in Galve et al. [2020b].

The GPU-PET reconstructor is a key element for the research presented in this thesis. In chapter 4, a modification of the projection-backprojection routine to produce a multi-ray scheme combined with a super-iterative algorithm that refines the data acquired to increase image resolution was presented. In section 5.4, we used the GPU-PET software with data from the clinical Biograph mMR scanner (section 2.2.2) to evaluate the scatter estimation derived from the Ultra-fast Monte Carlo PET (UMC-PET) simulator, a GPU-accelerated MC simulator that is presented in chapter 5. Finally, in chapter 6 we will use the GPU-PET reconstructor as the reference standard against a fully-integrated Monte Carlo based reconstruction with the UMC-PET simulator.

In next section 3.2 we described the details of the System Response Matrix (SRM), its implementation with the MAP-OSEM algorithm, the input details, and the GPU code implementation in the CUDA-Fortran platform. In section 3.3 we give the details of the simulations and PET data used, and in section 3.4 we show the reconstructed images of these data. We discuss about the methods and results in section 3.5.

3.2 Code description

The GPU-PET reconstructor is a fast reconstruction code for general PET scanner geometry. The scanner is coded in a cartesian coordinates system that allows us to reconstruct any geometry. We implemented the OSEM algorithm with Bayesian penalized likelihood in the NVIDIA CUDA platform (with the PGI CUDA Fortran Compiler) to accelerate the main routines of projection, convolution, and other image filters and operations. In this section we explain the main features of the GPU-PET reconstructor, the SRM, the penalized OSEM reconstruction algorithm, and the details of the inputs and its implementation.

3.2.1 The System Response Matrix and PET corrections

The SRM is one of the most important pieces of iterative reconstruction methods [Iriarte et al., 2016]. In the GPU-PET reconstructor, we have decomposed the SRM in the following components: the PSF, the forward/backward projector, normalization, attenuation, scatter, and randoms contributions. The projection equation in 1.14 can be modified in the following manner:

$$\tilde{Y} = V \bullet U \bullet P(G * X) + R + S, \tag{3.1}$$

where given N LORs and M voxels, $\tilde{Y} \in \mathbb{R}^N$ is the projected data given an image $X \in \mathbb{R}^M$, and the additive components $R, S \in \mathbb{R}^N$ represent respectively the random and scatter events. Respect to 1.14, we have explicitly defined the SRM A in terms of the matrices G, P, U, and N. $G \in \mathbb{R}^{M \times M}$ is the convolution kernel representing the PSF ($*$ is the convolution operator), $P \in \mathbb{R}^{N \times M}$ is the forward projection operator, $U \in \mathbb{R}^N$ are the attenuation factors, and $V \in \mathbb{R}^N$ are the normalization factors ($\bullet$ represents element-wise multiplication).

3.2.1.1 Point Spread Function modeling

The PSF models all the physical effects that deteriorate image quality (see section 1.4.2.1), and in a more general way, it defines the scanners intrinsic resolution. For fast image reconstruction, we chose to implement the PSF in the image domain [Reader et al., 2003] with a Gaussian kernel. The convolution is separately applied in the three axis,

$$\hat{X} = G * X = G_{\mathcal{Z}} * G_{\mathcal{Y}} * G_{\mathcal{X}} * X. \tag{3.2}$$

Here we show explicitly the kernel convolution for the $\mathcal{X}$ axis:

$$\begin{aligned} \hat{x}_{j_{\mathcal{X}},j_{\mathcal{Y}},j_{\mathcal{Z}}} &= \sum_{\Delta_{j_{\mathcal{X}}}=-N_{G_{\mathcal{X}}}}^{N_{G_{\mathcal{X}}}} x_{j_{\mathcal{X}}-\Delta_{j_{\mathcal{X}}},j_{\mathcal{Y}},j_{\mathcal{Z}}} C_{G_{\mathcal{X}}} exp\left(-\frac{\Delta_{j_{\mathcal{X}}}^2}{2\sigma_{\mathcal{X}}^2}\right) \\ C_{G_{\mathcal{X}}}^{-1} &= \sum_{\Delta_{j_{\mathcal{X}}}=-N_{G_{\mathcal{X}}}}^{N_{G_{\mathcal{X}}}} exp\left(-\frac{\Delta_{j_{\mathcal{X}}}^2}{2\sigma_{\mathcal{X}}^2}\right), \end{aligned} \tag{3.3}$$

where we perform each convolution kernel equivalently for every cartesian axis ($\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$). Note here that we have extended the array notation of $X \in \mathbb{R}^M$ to a three dimensional array $X \in \mathbb{R}^{M_{\mathcal{X}} \times M_{\mathcal{Y}} \times M_{\mathcal{Z}}}$ representing the three cartesian axis ($M = M_{\mathcal{X}} \times M_{\mathcal{Y}} \times M_{\mathcal{Z}}$). $N_{G_{\mathcal{X}}}$ is the the kernel size, $\sigma_{\mathcal{X}}$ is the Gaussian deviation, and $C_{G_{\mathcal{X}}}$ is a normalization constant for the kernel. This coordinate-based PSF allows us to reduce the integration neighbourhood in the total convolution from $N_{G_{\mathcal{X}}} \times N_{G_{\mathcal{Y}}} \times N_{G_{\mathcal{Z}}} \approx 15^3$ to $N_{G_{\mathcal{X}}} + N_{G_{\mathcal{Y}}} + N_{G_{\mathcal{Z}}} \approx 3 \times 15$ (the usual kernel size is around 15 voxels), thus reducing the computational burden of this step. It can be easily proben that this decomposition is equivalent to a 3D Gaussian blurring kernel (assuming the correlation between the three principal axes is zero).

In this work, we have estimated the Gaussian width from a simulation of a single point source at an arbitrary position (in our case, 1 cm off-center), reconstructed without the PSF kernel. The reconstructed image is equivalent to the ideal point source blurred with the intrinsic scanner resolution, so that the projection kernel gives the most accurate estimation of the data [Kotasidis et al., 2011, Reader et al., 2003]. Therefore, we can fit $\sigma_{\mathcal{X}}$, $\sigma_{\mathcal{Y}}$ and $\sigma_{\mathcal{Z}}$ to match the blurred image. The PSF can also be measured using the experimental resolution measurements of the scanner. In section 6.2.1 (chapter 6) we describe a modification of this routine to introduce spatially variant PSF using various measurements of point sources at difference locations. These extensions were not part of the author's work, and they are not presented in this manuscript in detail.

Note that the fixed PSF through the whole FOV is, in general, a rough approximation for real scanners, in which DOI effects degrade the resolution off-center (figure 1.9).

In the case of the scanners used in this chapter (mentioned below in 3.3), the phantoms used filled the center of the FOV where we can maintain this approximation, or additional DOI information was provided by the detector to alleviate PSF degradation.

3.2.1.2 Forward/backward projection scheme

The forward projector operator can be represented as a matrix $P \in \mathbb{R}^{N \times M}$. In the GPU-PET reconstructor, this matrix is computed on the flight using simple 3D discrete line integrals with Gaussian weights (similar to one of the approaches used in [Kang et al., 2021]). The algorithm computes the intersection between the given crystal coordinates and the scanner FOV, and it walks through the FOV with a fixed step of 0.9 times the pixel size to ensure good compromise between sampling and computation time. We accumulate the value of the closest voxel (nearest neighbour approximation) weighted by the Gaussian of the perpendicular distance with the line (FWHM equal to twice the voxel size) to reduce aliasing defects because of the sampling step. The projections are normalized to reproduce the physical length of the LOR times the image voxel values. In the backprojection, we used the same model, but we accumulate the projection values on the stepped voxels. In figure 3.1 a visual representation of the algorithm is shown. We see that there are over/under sampled voxels that have artificially mismatched sensitivity because of the nearest neighbour approximation. In practice the sensitivity will be averaged among the huge number of LORs stepping through each voxel, mitigating single LOR defects. We also have to account that the backprojection is performed with the same operator, thus compensating any deviation in the reconstruction.

3.2.1.3 Normalization

The normalization factor of every LOR multiplies the projection as stated in equation 3.1. The UMC-PET requires a calibration acquisition that must represent a cylindrical object of known dimensions and uniform activity. This calibration is compared with the projections of an equivalent numerical cylinder filling the whole FOV. Intrinsic detector efficiencies, geometrical efficiencies, and projection aliasing correction factors are implicit in this term. The normalization factor estimated is given by

$$v_i = \frac{q_i}{\tilde{q}_i} \frac{\tilde{l}_i}{l_i}, \tag{3.4}$$

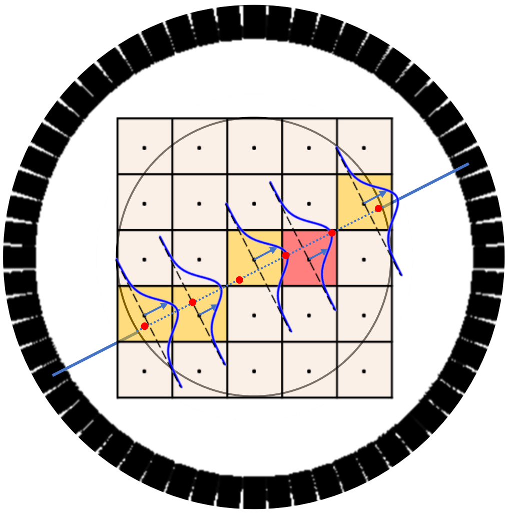


Figure 3.1: 2D scheme of the projection computation. The line integral performs discrete steps through the FOV (red dots), accumulating each pixel contribution weighted by the Gaussian of the distance pixel-LOR. The pixel color intensity represents its total contribution to the line integral (the darkest one has double contribution because of discrete sampling).

where v_i is the normalization factor of a LOR i, q_i is the measured acquisition for calibration, $\tilde{q}_i$ is the equivalent LOR projection, l_i is the theoretical LOR distance through the calibration phantom, and $\tilde{l}_i$ is the analytical LOR length through the FOV. The ratio $\tilde{l}_i/l_i$ is a projection length correction in the case the calibration phantom and the reconstruction FOV do not have the same size.

It is important to remark, that the calibration q_i should be precorrected for random events, scatter and attenuation, since the projection value $\tilde{q}_i$ does not account for these corrections. This approach assumes direct normalization with a cylindrical phantom of equal (o larger) size to the FOV [Theodorakis et al., 2013], but we can introduce component based normalization [Badawi and Marsden, 1999, Defrise et al., 1991, Hoffman et al., 1989] adapted to the FOV that we want to use, independently on the calibration phantom.

3.2.1.4 Attenuation

To estimate the attenuation coefficients we used the forward projection operator and PSF defined in the previous sections 3.2.1.2 and 3.2.1.1. The equation 1.28 is explicitly

$$U = exp\left(-P(G * \mathcal{U})\right), \tag{3.5}$$

where $\mathcal{U} \in \mathbb{R}^M$ is the µ-map for 511 keV photons, and the exponential operation is performed element-wise. For the µ-maps in HU units, as it is the usual case for CT µ-maps (MRI-derived µ-maps are given in linear attenuation units), we first applied the bilinear transformation of equation 5.7 [Burger et al., 2002, Nakamoto et al., 2002, Schneider et al., 2000]. The attenuation coefficients can also be provided from any external method.

3.2.1.5 Scatter and random correction

Both scatter and random estimations are additive quantities in the projection scheme, and neither of them is processed by the code (they should be facilitated by the scanner or estimated via external tools). Therefore they are treated in a similar way. The user defines in the general inputs (see section 3.2.3) whether to add them to the forward projection scheme (as explicitly stated in equation 3.1) or subtract them from the data (in this case, negative values are set to zero).

The UMC-PET simulator is a GPU-accelerated MC simulator that will be detailed in chapter 5. In this chapter, we have applied the UMC-PET simulator to estimate the scatter correction only for the acquisitions with prominent scatter contribution (in general this contribution is not relevant for small animals, see the results of scatter fraction on table 5.3 of section 5.3). The application of the UMC-PET simulator for scatter correction is detailed in section 5.4.

3.2.2 Basics of the OSEM algorithm

The reconstruction algorithm chosen for the GPU-PET reconstructor is the OSEM [Hudson and Larkin, 1994]. In section 1.4.2.3, we already gave the details about this reconstruction method. To recall the main features, OSEM is an accelerated version of the MLEM [Shepp and Vardi, 1982] based on data subsets, achieving an acceleration factor approximately equal to the number of subsets K. The maximum likelihood is optimized using the expectation maximization optimizer [Dempster et al., 1977]. Combining the model proposed in equation 3.1 with the OSEM iterative equation 1.21, each image update is computed as follows

$$X^{(n,k+1)} = X^{(n,k)} \frac{G * \Gamma_{S_k}^{(n,k)}}{G * \Lambda_{S_k}}, \tag{3.6}$$

where we have used matrix notation. $X^{(n,k)}$ is the image for the subset k at iteration n, and G is the image blurring kernel. We introduce here the correction factors image ($\Gamma_{S_k}^{(n,k)} \in \mathbb{R}^M$) and the sensitivity image ($\Lambda_{S_k} \in \mathbb{R}^M$). S_k refers to the set of LORs in the subset k. These images are defined as follows

$$\begin{aligned} \Gamma_{S_k}^{(n,k)} &= P_{S_k}^T U \bullet V \bullet \frac{Y}{U \bullet N \bullet P_{S_k}\left(G * X^{(n,k)}\right) + R + S} \\ \Lambda_{S_k} &= P_{S_k}^T (U \bullet V), \end{aligned} \tag{3.7}$$

where Y is the data acquired. G, P, U, V, R and S compose the SRM of equation 3.1 (P_{S_k} and $P_{S_k}^T$ are the projection and backprojection operators of a subset S_k), and the details of each component have been given in the previous section 3.2.1.

3.2.2.1 Bayesian priors with the OSEM algorithm

We have implemented two different prior functions in the GPU-PET reconstructor: the Median Root Prior (MRP) [Alenius and Ruotsalainen, 1997], and the Relative Differences Prior (RDP) [Nuyts et al., 2002]. In Alenius and Ruotsalainen [1997], the MAP-EM equation 1.25 with MRP is modified as follows

$$x_j^{(n,k+1),MAP} = \frac{x_j^{(n,k+1),OSEM}}{1 + \beta \frac{x_j^{(n,k),MAP} - M\left(x_j^{(n,k),MAP}\right)}{M\left(x_j^{(n,k),MAP}\right)}}, \tag{3.8}$$

where $x_j^{(n,k),OSEM}$ is the image after the OSEM update step (equation 3.6, note that capital letters were used for matrix notation, whereas lower case letters refer to the elements of a given matrix or vector), $x_j^{(n,k),MAP}$ is the image after the MAP update step, and $M(x_j)$ is the median in a neighbourhood of size $3 \times 3 \times 3$ centered in x_j. β is a hyperparameter that controls the regularization strength, and it must be in the range $\beta \in [0, 1]$. In this case, the prior function Q is not explicitly defined, though Bettinardi et al. [2002] defined its derivative *ad hoc*,

$$\left.\frac{\partial Q}{\partial x_j}\right|_{x_j^{(n,k)}} = \frac{x_j^{(n,k),MAP} - M\left(x_j^{(n,k),MAP}\right)}{M\left(x_j^{(n,k),MAP}\right)}. \tag{3.9}$$

The RDP prior defined in Ahn et al. [2015], Nuyts et al. [2002] is

$$Q_{RDP}(X) = \sum_{j} \sum_{j' \in N_j} w_{jj'} \frac{\left(x_j - x_{j'}\right)^2}{x_j + x_{j'} + \gamma \left|x_j + x_{j'}\right|}, \tag{3.10}$$

where $w_{jj'}$ are weights inversely proportional to the voxel distance (1, $1/\sqrt{2}$ and $1/\sqrt{3}$, in a neighbourhood of size $3 \times 3 \times 3$), N_j is the neighbourhood of voxel j, and γ controls the relative difference penalization. The derivative of this equation is given by:

$$\left.\frac{\partial Q_{RDP}}{\partial x_j}\right|_{x_j^{(n,k)}} = \sum_{j' \in N_j} w_{jj'} \frac{\left(x_j - x_{j'}\right)\left(\gamma \left|x_j - x_{j'}\right| + x_j + 3x_{j'}\right)}{\left(x_j + x_{j'} + \gamma \left|x_j + x_{j'}\right|\right)^2} \tag{3.11}$$

3.2.3 Code inputs

We have designed the GPU-PET inputs to ensure great adaptability of the code to any scanner geometry. We have left most of the parameters to the user control to optimize the reconstruction, and the scanner geometry is coded in two simple files (crystal coordinates file, and LOR-to-crystals file).

- General parameters

  In a general file, we describe all the variables that control the parameters of the reconstruction algorithm. There are image parameters, such as the number of voxels (in the three dimensions) or the FOV size (both transaxial and axial), other parameters defining the OSEM performance (number of iterations, number of subsets, the PSF Gaussian width, the sampling Gaussian width in the projection routine, and the β and γ values for the priors), or the total number of LORs (this number must suit the system employed). We also include here the number of threads and blocks used for the forward/backward projection subroutine.

- LOR histogram files

  Here we find the data file of the acquisition, the calibration (to compute the normalization), the attenuation factors, and the scatter and random estimation. These files use LOR histogram format [Kadrmas, 2004] with no predefined ordering. We use floating point numbers with single precision.

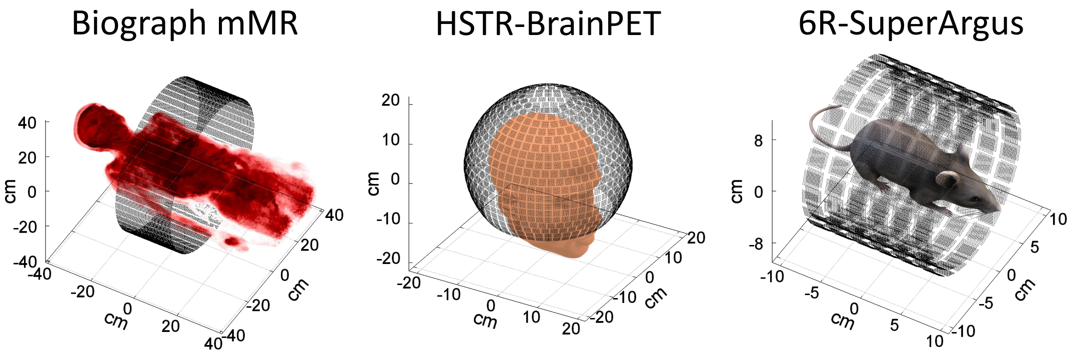


Figure 3.2: From left to right, the 3D plot of the crystal coordinates for the Biograph mMR scanner (section 2.2.2), the HSTR-BrainPET scanner (section 3.3.1), and the preclinical 6R-SuperArgus scanner (section 2.2.1). Different objects have been inserted to ease 3D perspective.

- Crystal coordinates

  This is one of the key files to ensure the adaptability to any scanner geometry. We define all the projections from two points in the cartesian space for both scanner crystals defining the LOR. This file defines the cartesian coordinates $\vec{r}_c$ of every single crystal c. A representation of the crystal coordinates file of three different scanners is shown in figure 3.2. The image will be centered in the center of the cartesian system employed. In this file we write the crystal coordinates in an arbitrary order.

- LOR index to coordinates

  In this file, we define a matrix relating every LOR index i with two crystals from the coorcinates file, $(c_{i,1}, c_{i,2})$, so that the LOR spatial location is defined by the crystal coordinates file $(\vec{r}_{c_{i,1}}, \vec{r}_{c_{i,2}})$. We use short integers (2 bytes) unless the size of the scanner requires a larger number of crystals.

3.2.4 GPU implementation

We have implemented the main routines of the GPU-PET reconstructor in the GPU to reduce the computation time. We can distinguish the forward/backward projection routine, which consumes most of the time of the reconstruction, and the image filters that apply the PSF blurring and the regularization step.

3.2.4.1 Forward/backward projection routine

We perform both the forward projection and backward (FW/BW) projection (also called backprojection) in the same routine. This implies the computation of the image of correction factors and the sensitivity images in equation 3.7 (the sensitivity image is computed only once). In practice, we perform this step in several routine calls. Prior to this whole step, the image to be projected is loaded on the GPU global memory and the output images are initialized to zero. The GPU routine works with a structure of blocks and threads, having $N_{threads}$ and N_{blocks} defined by the user. In every call, we work with $N_{threads} \times N_{blocks}$ LORs, and we repeat the routine until all the LORs in the subset are done. Prior to every single call we load all the correspondent LOR values in the global memory (the data acquisition, the calibration acquisition, the attenuation factors, the scatter estimation, the random estimation, and the LOR coordinates). In the first lines of the routine, we load these input values from the global memory to the thread local memory to reduce computational time in repetitive access to the variables. The forward and backward projection for every LOR are computed sequentially with the model detailed in section 3.2.1.2. The backward projection uses atomic instructions since we may find overlapping of different LORs (thus threads) on the same voxel. We have verified that this does not represent a significant performance penalty since it is very unlikely to find multiple threads accessing the same voxel at the same time. The main time consuming operations are global memory accesses to the projected image.

It is important to remark that the MLEM equation 3.6 does not require to FW/BW project the LORs with no counts. For this reason, prior to the reconstruction we select the positive LORs in the acquisition to estimate the correction image. Since the imaged objects may not fill the whole FOV, and PET limited statistics, this reduces considerably the number of active LORs. The sensitivity image, in contrast, should account for the whole data set. Fortunately, we just need to compute it once and for all in a given scanner. We have separated the computation of both images in the code.

3.2.4.2 Image filters

We have defined the GPU accelerated routines that work on the image domain with a similar voxel-wise structure. In the GPU we define a 3D grid of blocks and threads matching the image size, so that every thread works in parallel over an output voxel. Input and output images are kept in global memory. These routines are:

- PSF blurring kernel. Actually three sequential routines for the three cartesian axis $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ (see 3.2.1.1). The convolution kernels are precomputed. At run-time the sum in equation 3.3 is performed in parallel for each voxel.
- Regularization kernels. Both the MRP and RDP are applied as post-processing filters of the OSEM image. The only input is the precomputed OSEM image. For explicit details of the operations see section 3.2.2.1.
- Image update. OSEM update of every voxel, given the last iteration, and the correction and sensitivity images computed in the forward/backward routine.

3.3 PET data acquisitions and simulations

We have ran simulations of the Biograph mMR scanner (section 2.2.2) and the HSTR-BrainPET scanner (details in next section 3.3.1). Real acquisitions of the 6R-SuperArgus preclinical scanner are also shown. In chapter 5.4 we show real acquisitions reconstructions of the Biograph mMR with the GPU-PET reconstructor and scatter estimation from the UMC-PET simulator.

3.3.1 Simulations of the Biograph mMR scanner and the HSTR-BrainPET scanner

We have ran simulations of both the Biograph mMR (see section 2.2.2) and the HSRT-BrainPET scanner (see 2.2.3). In the case of the HSTR-BrainPET simulations, we compared the results with 26 mm and 9 mm crystal depth to assess the DOI effect. In this case, the simulated activity was corrected to have the same equivalent noise as for the thicker crystals. The energy window implemented (for both scanners) is 400-700 keV. We have used a Derenzo phantom (see section 2.5.1, left-hand side of figure 2.8), a Cologne phantom (section 2.5.2, figure 2.9), and a Zubal phantom (section 2.5.3, figure 2.10). In the case of the Zubal phantom, the noise was measured as the percentage ratio between the standard deviation and mean activity inside the elliptic region drawn on figure 2.10 (similar to equation 2.1).

We ran all the simulations in the UMC-PET simulator (see chapter 5). We used $1{\cdot}10^9$ decays of ^{18}F-FDG, equivalent to an acquisition of 10 minutes with 1.78 MBq (0.465 µCi). To ease the resolution comparison between the Biograph mMR and the HSTR-BrainPET scanner caused by the differences in the geometry and crystals size of each scanner, the simulations did not include an attenuation object (no attenuation and scatter estimations were needed for the reconstruction). We also simulated an homogeneous cylinder filling the whole FOV (25 cm diameter and 20 cm length in the HSTR-BrainPET, and 60 cm diameter and 25 cm length in the mMR) to estimate the calibration factors.

3.3.2 Acquisitions with the preclinical PET 6R-SuperArgus scanner

All the real acquisitions of this chapter were obtained with the 6R-SuperArgus scanner presented in section 2.2.1. Acquisitions were done at the Laboratory of Medical Imaging of Hospital Gregorio Marañón (Madrid, Spain). All experimental procedures have been done in compliance with the European Communities Council Directive 2010/63/EU and submitted for approval by the Institutional Animal Care and Use and Ethics Committee of the Hospital General Universitario Gregorio Marañón (HGUGM), supervised by the Comunidad de Madrid according to the Annex X of the RD 53/2013. The subjects were kept at the animal housing facilities of the Unidad de Medicina y Cirugía Experimental (UMCE-HGUGM) in Madrid, Spain.

We show a couple of examples on phantoms, followed by three studies with live small animals. We reconstructed all the acquisitions with the randoms corrections supported by the scanner (subtracted to the data set), in the energy window of 425-600 keV.

The Derenzo phantom detailed in section 2.4 has been acquired with hot and cold rods. The hot Derenzo was filled with 250 µCi of ^{18}F-FDG (9.25 MBq) measured with a well-counter (accuracy of ±5%), and $6.19{\cdot}10^8$ counts were detected. The cold Derenzo phantom has a background activity of 400 µCi of ^{18}F-FDG (14.8 MBq) and $2.55{\cdot}10^9$ counts. Both acquisitions were one hour long.

The first animal acquisition shows a cardiac study on a mouse of 15 g injected with 220 µCi ^{18}F-FDG (8.14 MBq), 30 minutes after the injection. The acquisition took 10 minutes and it had $9.60{\cdot}10^7$ coincidences which have been gated in 8 steps of the rat heart beat.

We show a multi-animal study with four mouse injected with 0.4 µCi of ^{18}F-FDG each (148 kBq), acquired simultaneously for 10 minutes. A total of $4.40{\cdot}10^8$ coincidences were detected.

A Wistar female rat of 255 g injected with 500 µCi of ^{18}F-FDG (18.5 MBq) was acquired for 15 minutes, 40 minutes after injection. A total of $1.63{\cdot}10^8$ coincidences were detected. This acquisition was part of an awake study with rigid motion correction [Arias-Valcayo et al., 2019]. Four hot point sources of 7 µCi were attached to the head to identify the motion, and they can be observed in the reconstruction. The images shown in this work correspond to the reference acquisition with the animal anaesthetised.

3.4 Results

3.4.1 Simulations of the Biograph mMR scanner and the HSTR-BrainPET scanner

In all the reconstructions of this section (simulation details in section 3.3.1) the PSF used for the HSTR-BrainPET was 1.75 mm FWHM in the three directions, and 4 mm FWHM for the Biograph mMR. In figures 3.3 and 3.4 we show the reconstruction of the Derenzo phantom and the Cologne phantom. In both cases the voxel size used is $0.4{\times}0.4{\times}2.05$ mm^3 for the HSTR-BrainPET and $0.5{\times}0.5{\times}2.05$ mm^3 for the Biograph mMR. In the HSTR-BrainPET we can clearly see the sources up to 2 mm diameter (for the 9 mm length detectors), whereas for the Biograph mMR we can hardly see the 3 mm rods of the Derenzo phantom.

In figure 3.5, we show the results of the Zubal phantom. The voxel size used is $0.5{\times}0.5{\times}2.05$ mm^3 in both scanners. In this case, we have adjusted the number of iterations to compare the image quality obtained in both scanners at an equivalent noise level. The improvement in image quality is clearly seen.

3.4.2 Preclinical PET 6R-SuperArgus scanner

Figures 3.6, 3.7 and 3.8 (hot and cold Derenzo phantoms and mouse cardiac study) have a voxel size of $0.25{\times}0.25{\times}0.81$ mm^3, and figures 3.9 and 3.10 (four simultaneous mice and rat acquisitions) used $0.35{\times}0.35{\times}0.81$ mm^3. The number of iterations and regularization parameters have been determined from experimental trials, comparing the images at different number of iterations and β values to assess the regularization

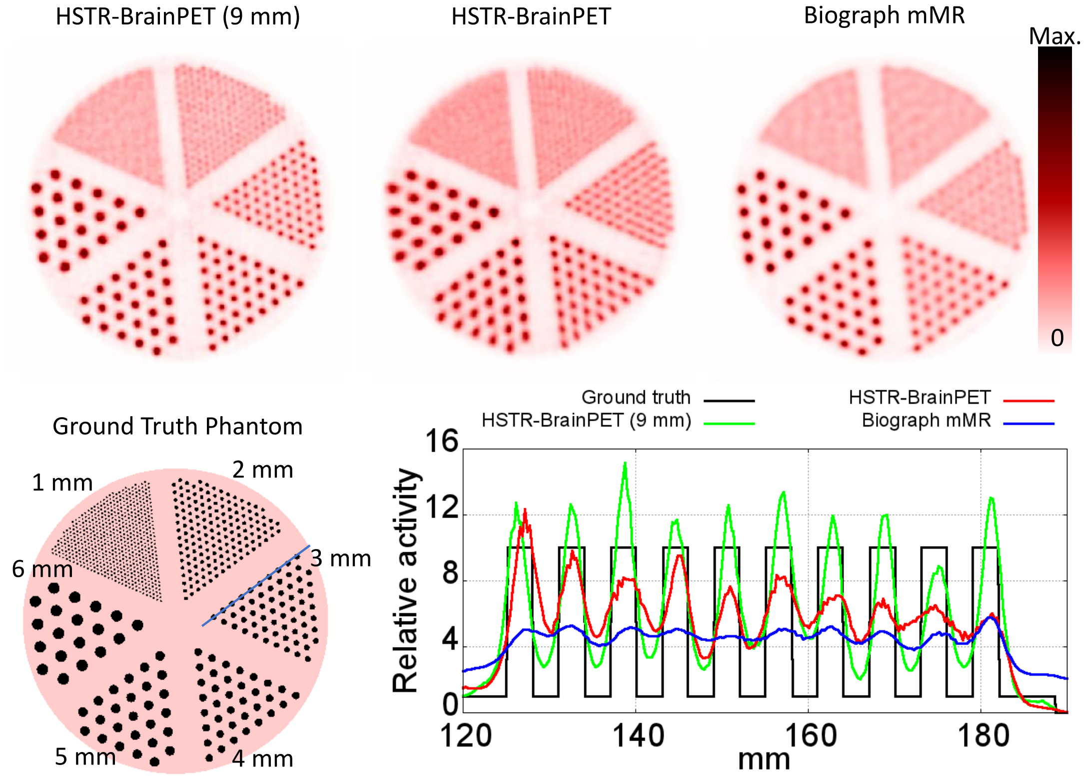


Figure 3.3: Top row, from left to right, tranversal view of a single slice of the Derenzo phantom in the HSTR-BrainPET scanner with 9 mm depth (left) and with 26 mm depth (center), and in the Biograph mMR scanner (right). All the images have 4 iterations and 20 subsets, no regularization was used. Bottom row, left, we show the simulated Derenzo phantom (rods to background ratio 10:1) with the rods diameters of each section. Bottom-right, we show a profile through the 3 mm rods (blue line in the ground truth phantom).

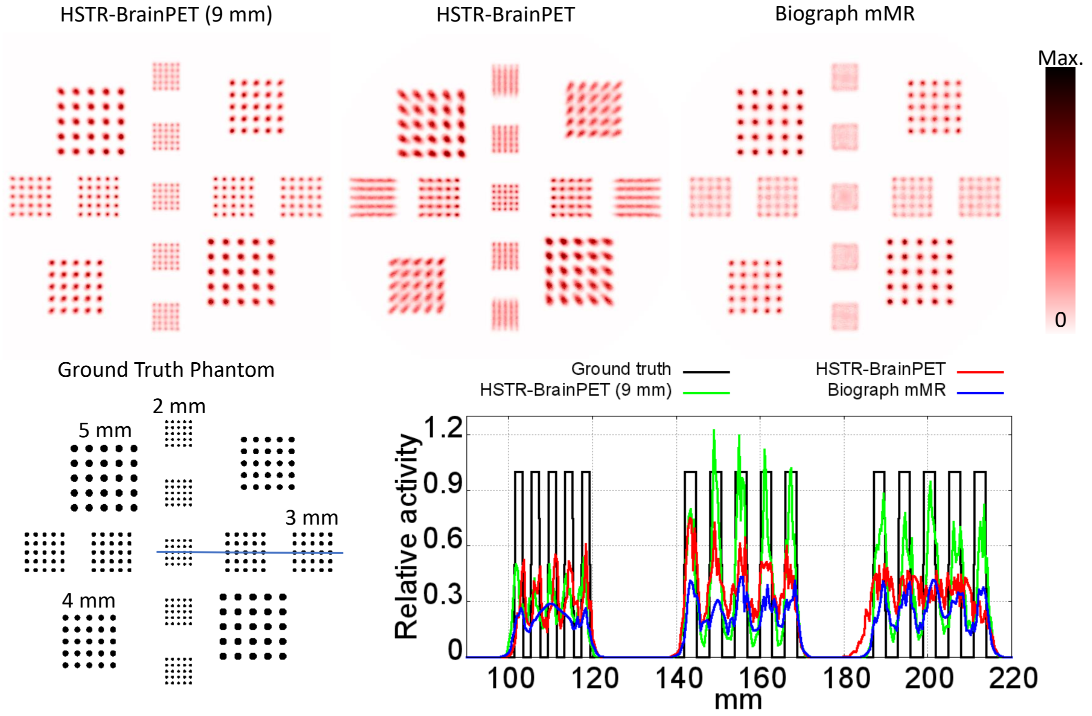


Figure 3.4: Top row, from left to right, tranverse view of the Cologne phantom in the HSTR-BrainPET scanner with 9 mm depth (left) and with 26 mm depth (center), and in the Biograph mMR scanner (right). All the images have 4 iterations and 20 subsets, no regularization was used. Bottom row, left, we show the simulated Cologne phantom with the sources diameter of each square. Bottom-right, we show a profile through the 2 mm and 3 mm sources (blue line in the ground truth phantom).

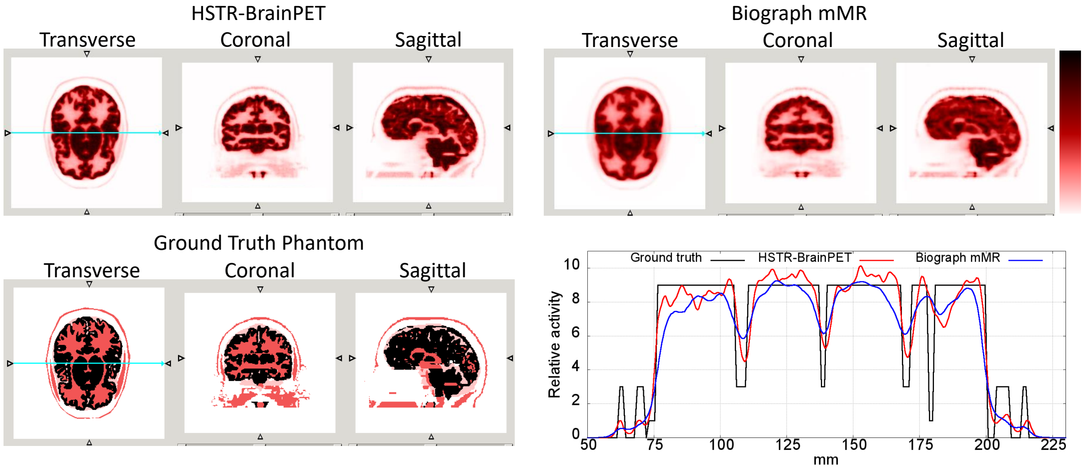


Figure 3.5: Top row: transverse, coronal and sagittal views of the Zubal phantom reconstructed in the HSTR-BrainPET scanner and the Biograph mMR scanner. We used 3 iterations of 20 subsets for the HSTR-BrainPET (4.24% noise), and 1 iteration of 20 subsets for the mMR (4.61% noise). The number of iterations was adjusted to compare equivalent noise levels. The noise was measured using a uniform region in the cerebellum (see figure 2.10). Bottom row: (left) the ground truth Zubal phantom, (right) line profile through the transverse views.

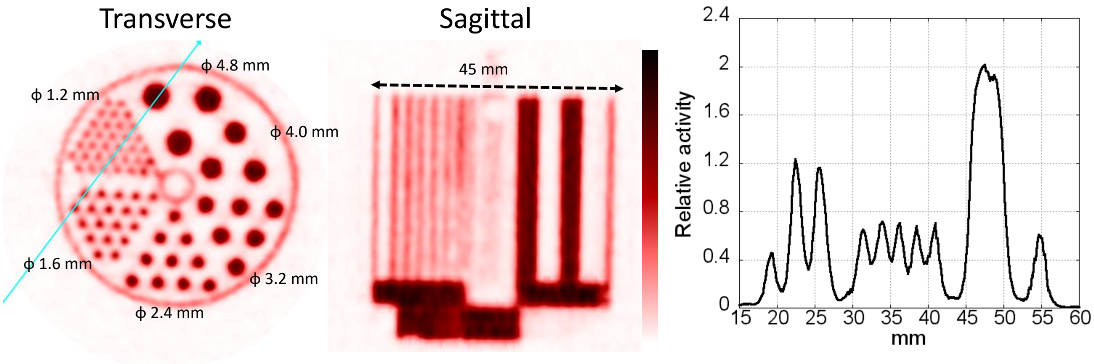


Figure 3.6: Transverse and sagittal planes of a single slice of the hot Derenzo phantom in the 6R-SuperArgus scanner reconstructed with 25 iterations and 6 subsets. The image regularization uses RDP with β=0.01 and γ=2. In the right-hand side of the image we show the line profile of the blue arrow through the transverse plane. The rods diameter (1.2 mm, 1.6 mm, 2.4 mm, 3.2 mm, 4.0 mm, and 4.8 mm) of each section of the phantom are shown in the transverse plane, and the total phantom diameter is shown in the sagittal plane.

strength for both priors. We found reasonable results for $\beta \in (0.1, 0.01)$, and $\gamma = 2$ for the particular case of the RDP.

In figures 3.6 and 3.7 we show the reconstructed images of the hot and cold Derenzo phantoms. In both cases a line profile is shown through the smaller rods of 1.2 mm diameter. In the case of the cold phantom of figure 3.7, we can notice that the phantom was not totally filled during the acquisition.

In figure 3.8 we show the reconstruction of the total acquisition for the mouse cardiac study in the 6R-SuperArgus scanner together with the evolution of the cardiac beat through the eight gates of the acquisition.

Figure 3.9 shows the simultaneous acquisition of four mice. Mice hearts' are visible around 3.5 cm off-center, showing submillimeter resolution across the whole FOV.

In figure 3.10 we show a rat head acquisition originally designed for a motion correction study. We can see two hot point sources attached to the rat head. The main structures of the brain (cortex, caudate, olfactory bulb, and cerebellum) have been marked in the three planes of the image.

In table 3.1 we summarize reconstruction times of the main routines of GPU-PET. The total time of each image is given by the initialization time (around 60 seconds for input load, subsets and kernels preparation), the computation of the sensitivity image,

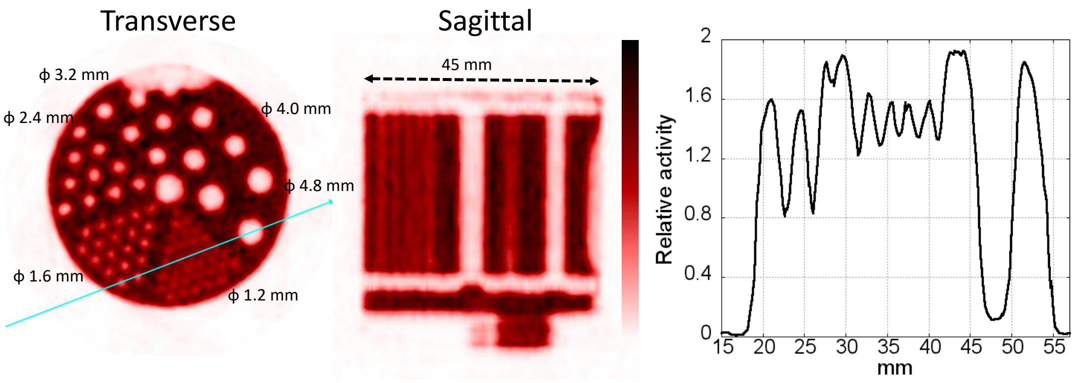


Figure 3.7: Transverse and sagittal planes of a single slice of the cold Derenzo phantom in the 6R-SuperArgus scanner reconstructed with 30 iterations and 6 subsets. Note the phantom was incompletely filled, as it can be observed in the upper side of the transverse plane. The image regularization uses RDP with β=0.01 and γ=2. In the right-hand side of the image we show the line profile of the blue arrow through the transverse plane. The rods diameter (1.2 mm, 1.6 mm, 2.4 mm, 3.2 mm, 4.0 mm, and 4.8 mm) of each section of the phantom are shown in the transverse plane, and the total phantom diameter is shown in the sagittal plane.

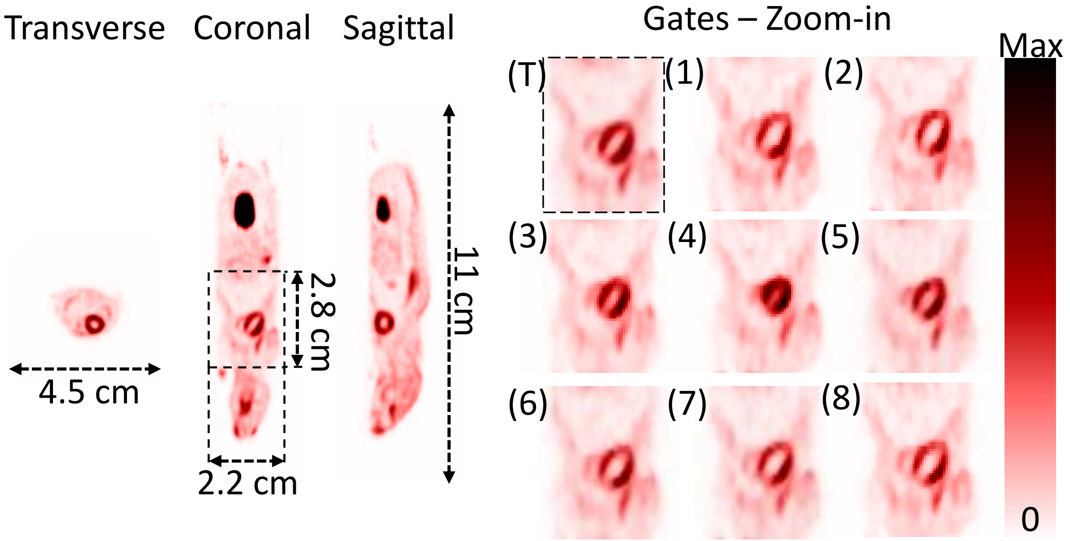


Figure 3.8: (Left) The three planes of the total acquisition of a mouse cardiac study in the 6R-SuperArgus scanner. (Right) Zoom-in to the heart area in the coronal plane with the total acquisition (T) and the eight separate gates of the cardiac beat. The color scale has been set to the 30% of the maximum value (in the bladder). We used 10 iterations of 6 subsets, with an MRP regularization with β=0.1 for the gated images to limit noise.

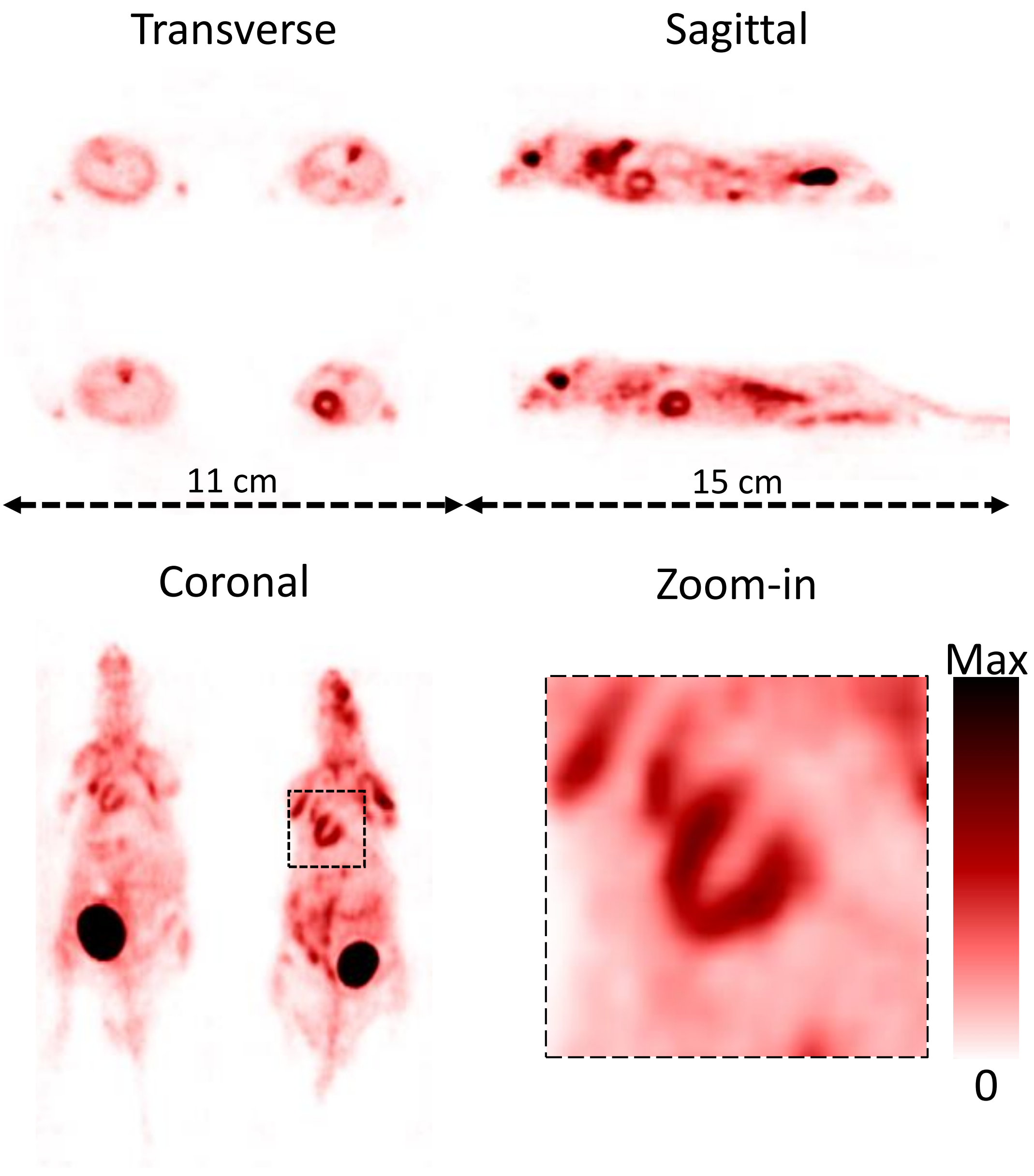


Figure 3.9: Transverse (up-left), sagittal (up-right) and coronal (down-left) planes of a simultaneous acquisition of four mice in the 6R-SuperArgus scanner. (Down-right) Coronal zoom-in insert of the heart of one of the mice. The color scale has been set to the 8% of the maximum activity. We used 10 iterations with 6 subsets, with RDP regularization with β=0.05 and γ=2.

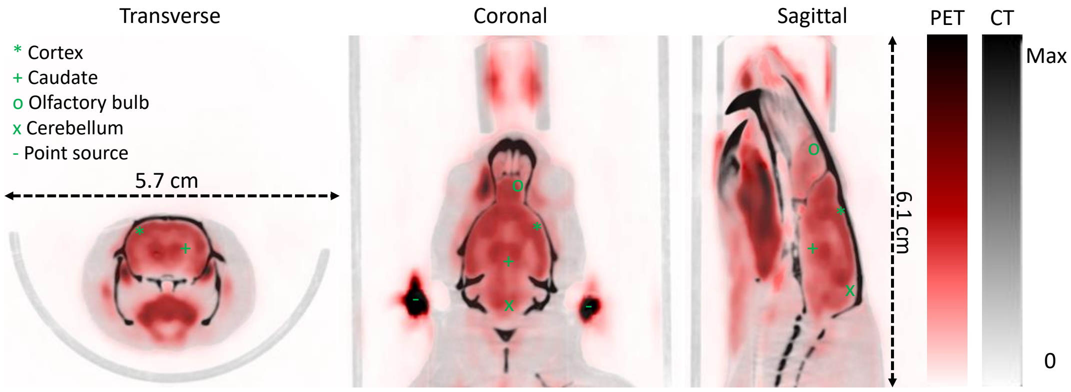


Figure 3.10: Rat head study with registered CT image in the 6R-SuperArgus scanner. The color scale has been set to a 5% of the maximum activity in the point sources, in order to properly see the brain structures. The legend in the left-hand side of the transverse panel shows the markdown for the cortex, the caudate, the olfactory bulb, the cerebellum, and the point sources attached to the head (originally attached for a rigid motion correction study). We used 10 iterations with 6 subsets and MRP regularization β=0.1.

and the computation of all the iterations. We find out that the computation of the sensitivity image takes much longer than an average iteration, thus in general we will precompute it. The FW/BW routine accounts for most of the computing time of the reconstruction (more than 80 % in all the cases). We can reduce the reconstruction time by increasing the number of subsets. In that case, the contribution of the image filters to the reconstruction time will increase, since calls to these routines are performed per subset. In general, larger objects need longer reconstruction times since they have larger number of active LORs (more FW/BW projection calls), and larger number of voxels also increases the computational time of all the routines.

3.5 Discussion

The GPU-PET reconstructor is a fast and flexible reconstruction software accelerated with GPU. The main novelty of this software is its scanner and data format description. The general data format (crystal coordinates and LOR-to-crystals matrix) presented is easily configured to any scanner geometry. This framework is a generalization, but not a limitation, from traditional sinogram format. In this chapter, we have successfully reconstructed two cylindrical scanners (the preclinical 6R-SuperArugs, and the clinical

Execution time in the GPU-PET reconstructor

Data	Image size	Sens. im. (s)	Av. it. (s)	FW/BW projection			Image filters			Total
				1 call (s)	(s)	(%)	1 call (s)	(s)	(%)	(s)
Derenzo	241×241×195	143	14.5	$2.02 \cdot 10^{-2}$	12.9	89.1	$9.17 \cdot 10^{-2}$	0.550	3.80	558
Cold Der.	241×241×195	116	13.3	$1.65 \cdot 10^{-2}$	11.7	87.9	$7.29 \cdot 10^{-2}$	0.437	3.29	561
Cardiac	241×241×195	114	4.56	$2.56 \cdot 10^{-2}$	3.66	80.3	$7.56 \cdot 10^{-2}$	0.454	9.95	216
Multianimal	343×343×195	178	31.3	$3.72 \cdot 10^{-2}$	29.3	93.5	$1.17 \cdot 10^{-1}$	0.700	2.23	1149
Rat head	241×241×195	116	4.82	$2.61 \cdot 10^{-2}$	3.90	80.8	$9.98 \cdot 10^{-2}$	0.599	12.4	214

Table 3.1: Computational time of the GPU-PET reconstructor for all the reconstructed images in the 6R-SuperArgus scanner. We have separated the results for the sensitivity image (Sens. im.) and average iteration (Av. it.). We show FW/BW routine and image filters (PSF blurring, regularization, and voxel update) times for an average iteration. For each routine, we show the time for each individual call, the total time per iteration, and the percentage time per iteration. The remaining of the time up to the total goes into general calculations in the CPU). In the case of the FW/BW projection, each call represents 512×512 LORs ($n_{threads} \times n_{blocks}$), and image filters are computed six times per iteration, one per each subset. Times correspond to an Intel(R) Xeon(R) W-2155 CPU @ 3.30GHz CPU host unit with 512 GB of RAM, and a 24 GB GeForce RTX 3090 with 10496 cores GPU.

Biograph mMR), and a spherical scanner (the HSTR-BrainPET). In chapter 5.4, we show further results of the reconstruction of sinogram data from real acquisitions of the Biograph mMR scanner.

The code is also very flexible since its modular structure allows for simple modifications to optimize its performance. In Galve et al. [2018], we explored the robustness of the reconstructions against hypothetical detector failure and on purpose removal of a section of the scanner to ease access to the subject inside the scanner. In Galve et al. [2021c] (chapter 4), we have modified the projection scheme to apply a novel superiterative algorithm for resolution improvement. In Arias-Valcayo et al. [2019, 2020], the LOR histogram format was modified to accept list-mode data. For the list-mode reconstruction, the code performs fast enough in common GPUs to allow for a full 3D OSEM iterative reconstructions in real time (from table 3.1, an average of $\sim 1 \cdot 10^7$ LORs/s are processed in the FW/BW routine), achieving unprecedented high quality real-time PET imaging [Udias et al., 2018]. A version of the software described here is currently under use in the SuperArgus family of PET scanners.

A detailed comparison of the GPU-PET performance against other reconstruction software or CPU implementation is beyond the scope of this thesis. Nonetheless, in the literature we may find a wide range of GPU-accelerated image reconstructors. Meng et al. [2019] reported a 113× acceleration factor compared with CPU-based calculations using a LOR driven approach for the projection, and separate voxel-driven approach for the backprojection. This model was based on a precomputed SRM for a dual-head PET scanner, not comparable in size with any of the scanners presented in this chapter. Zhou and Qi [2014] reported a 3.5× acceleration factor in the projection (not considering image blurring) for list-mode data, compared with a 16 CPU cores working in parallel. In this case, the scanner used is of equivalent size to the Biograph mMR and the SRM factorization is similar to the one used here. The projection performance, however, is hard to compare since we do not know how many events they are accounting in their timing results. In terms of total reconstruction time, they reported a maximum 27.7 s per iteration for the non-TOF projection kernel on the GPU on an image of 300×300×89 voxels (bigger than the times we have shown in table 3.1). For the image blurring kernel, they measured 95 ms, whereas we measured 16 ms for the 241×241×195 images and 24 ms for the 343×343×195 image (this represents the partial time of a single blurring call in the column "Image filters" of table 3.1). It is important to remark that the GPU device employed by Zhou and Qi [2014] (NVIDIA GTX480) is not comparable with current GPUs (such as the GeForce RTX 3090 used in our work). Other authors who reported improvements comparing CPU computing with GPU computing were Herraiz et al. [2011] (72×), Pratx et al. [2009] (50×), or Barker et al. [2009] (40×).

The image blurring PSF has been optimized for speed. The PSF estimation can be performed in several ways, but we preferred to use MC estimated PSFs to avoid error propagation of further corrections, such as normalization, attenuation, scatter, or random correction, which must be accounted when experimental approaches are employed to determine the PSF [Alessio et al., 2005, Kotasidis et al., 2011, Rapisarda et al., 2010]. It is important to remark that the PSF kernel, as it is detailed in section 3.2.1.1, does not account for the heterogeneous resolution along the FOV, which is a poor approximation for the majority of scanners. We have explored spatially-variant PSF (SV-PSF), achieving homogeneous resolution without losing computational performance, but these studies were not part of the author's PhD. Thus, they are not presented in this manuscript in detail. More complex SRM schemes may introduce tuves of response (TORs) or explicit MC based representations of the SRM, among others (see section 1.4.2.1). In chapter 6, we compare the SV-PSF against a fully MC projection method. Even though the image blurring implementation is very flexible, some authors recommend the use of PSF based reconstructions cautiously since it may yield Gibbs artifacts and overestimation of quantitative lesion estimates [Munk et al., 2017, Rogasch et al., 2021, 2014]. We did not observe any artifacts in the images we have shown for the 6R-SuperArgus scanner, and in general we have verified that such artifacts only result if using too wide PSF together with over-iterated images. In figure 3.10, streak artifacts appeared around the point sources, but it is important to remark that the activity in the sources is much higher than the activity in the rat brain. This artifacts are seen because the maximum of the color scale has been set to a 5% of the maximum of the activity in the image (that is, the point sources maximum activity). In the case of the reconstructions of the HSTR-BrainPET (9 mm) of figure 3.3, we assume the overshoot observed in the profiles is caused by an inaccurate PSF estimation.

We developed the projector presented in section 3.2.1.2 to simplify the projection step, reducing the computational burden that other algorithms, such as the Siddon algorithm [Siddon, 1985], convey. Even though the algorithm exhibit simplified sampling, we have ensured that this is correctly compensated during the backprojection step, thus no artifacts are present in the images.

In section 3.2.2.1, we have given the details of two prior functions (the MRP [Alenius and Ruotsalainen, 1997, Bettinardi et al., 2002] and the RDP [Ahn et al., 2015, Nuyts et al., 2002]). The MAP-EM reconstruction (equation 1.25, [Green, 1990]) gives a clear insight on how to combine the priors with the OSEM algorithm, but the appropriate hyperparameter selection is not clear. Optimal values depend on the acquisition and noise level. This is one of the limitations of the method, since it requires a selection based on experimental trials comparing the images at different number of iterations and β values.

In figures 3.3 and 3.4 we have observed the image quality deterioration due to DOI in thick detectors. This problem might be solved with multilayered detectors that reduce the parallax error. In these images we can also appreciate that the HSTR-BrainPET scanner outperforms the image resolution obtained by the Biograph mMR. This is due to the reduced scanner diameter (32 cm in the HSTR-BrainPET, and 60 cm in the

Biograph mMR), that reduces the scanner non-colinearity effect, and the smaller crystal pitch (2 mm for the HSTR-BrainPET, and 4 mm for the Biograph mMR).

Chapter 4

Improved Image Reconstruction with Super-iterations

4.1 Introduction and motivation

One of PET's key disadvantages over other imaging modalities, such as CT or MRI, is its relatively low spatial resolution. PET resolution is produced by a number of different causes that are inherent to the technique (such as crystal size, positron range, photon non-collinearity, or depth of interaction) [Cherry et al., 2012, Moses, 2011]. In the next paragraphs we describe a variety of techniques proposed to improve spatial resolution. We focus on hardware-based solutions, super-resolution techniques, and software based-solutions.

One of the hardware-dependent issues that limits PET resolution is the width of the crystals in a detector block (∼4 mm in clinical scanners [Delso et al., 2011, Jakoby et al., 2007, 2009] and ∼1.5 mm in preclinical scanners [Nagy et al., 2013, Perez-Benito et al., 2018, Qi et al., 2011, Wang et al., 2006, Yang et al., 2019]). The resolution of PET detectors is also affected by the length of the scintillation crystals: larger crystals improve scanner sensitivity but degrade resolution due to depth-of-interaction (DOI) uncertainty [Cherry et al., 2012]. Some manufacturers have employed DOI uncertainty reduction solutions, although these usually necessitate complicated detectors and circuitry [Abbaszadeh et al., 2018, Mohammadi et al., 2017, Wang et al., 2006]. In general, hardware upgrades are limited by large costs.

Super-resolution (SR) are acquisition-based methods that enhance the sampling of the image space to improve resolution. Several acquisitions of the same object in different places, shifted by subpixel (or sub-crystal) distances, are obtained using controlled object or scanner motion. The set of low-resolution images can then be integrated in the image domain [Kennedy et al., 2006], the sinogram domain [Jeong et al., 2011, 2013], or during reconstruction [Li et al., 2014]. In each of these cases, SR approaches attempt to address the issue of limited sampling in the projection space by explicitly augmenting sampling through controlled displacements.

The use of software-based approaches is an inexpensive choice to improve image resolution. It has been demonstrated that using a realistic model in the image reconstruction process can compensate for these effects and increase image quality [Iriarte et al., 2016]. Analytical formulas, numerical integration methods [Huesman et al., 2000, Moehrs et al., 2008, Zeng et al., 1991], empirical kernels [Gong et al., 2017, Kotasidis et al., 2011, Tohme and Qi, 2009], or Monte Carlo simulations [Herraiz et al., 2006a, Iriarte et al., 2016] can be used to represent physical parameters related to the emission, transport, and detection of radiation.

In this chapter, we propose to further improve the image resolution with respect to the MAP-OSEM algorithm implemented in the GPU-PET reconstructor detailed in the previous chapter 3, although the algorithm is compatible with any reconstruction method. The presented approach iteratively refines the PET data into a new dataset with improved sampling. We were motivated by the finite crystal pitch size of the scintillator units forming the detector blocks, which limits the discrete LOR sampling in the projection space. With a LOR sampling redistribution strategy, our novel method addresses this problem without needing any hardware investment in motion controlled devices. We divide each LOR into parallel LOR subdivisions (subLORs) based on the subLOR projections of a converged image of the acquisition. Any standard image reconstruction process requires a series of implicit or explicit interpolations due to the discrete sampling of both the image and the projection space. Unlike sinogram-filtering methods [Herraiz et al., 2006b] or interpolated sinograms [Herraiz et al., 2008, Karp et al., 1988, Tuna et al., 2009], we use a reconstructed image of the data to estimate the improved sampling while preserving the statistics inside each original LOR (figure 4.1). Indeed, improving the data sampling with LOR subdivisions based on maximum-likelihood weights, effectively provides data-driven LOR sampling redistribution, which holds the potential to achieve better resolution than standard methods.

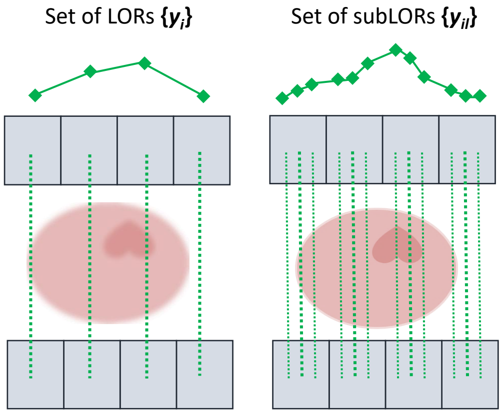


Figure 4.1: Schematic view of the transversal division of the LORs into three different subLORs. The resolution of the projection profile in the left-hand side image is limited by the crystal pitch. In the right-hand side, the resolution in the projection space is improved, which will be propagated to the reconstructed image.

Other LOR subdivisions methods have been proposed. [Kim et al., 2018] suggested to project fictitious subdivisions of the crystal units over a virtual ring-shaped detector, enhancing image resolution, although they used a uniform distribution among the sub-crystals. In [Jin et al., 2012], the authors implemented a distribution based on the normalization efficiencies to generate probabilistic LORs (pLORs) that expanded the sinogram axial compression (the span). Up-sampled sinograms from low resolution data can also be obtained with DL methods [Hong et al., 2018]. However, none of these approaches suggested the use of projections of a reconstructed image to generate the subdivisions.

The method was already applied to identify inter-crystal scatter events [Lee et al., 2018], recover triple coincidences in PET [Lage et al., 2015], and demultiplex data in pinhole SPECT [Moore et al., 2015]. In [Galve et al., 2017], we originally proposed to subdivide the crystal units of the detectors. In this chapter we summarize the work done on [Galve et al., 2021c], that was partially presented in [Galve et al., 2019a,b], a generalized approach based on the PSF resolution model.

This chapter is organized as follows. Section 4.2 explains the details of the algorithm. In section 4.3 we present a brief 2D example of the implementation of the algorithm. In section 4.4 we present the details about the acquisitions used, and the reconstruction software. In the next section, we evaluate the performance of the algorithm with an IQ phantom, a cold Derenzo phantom, and a mouse heart. Finally, we discuss about the work presented, technical disadvantages and possible improvements in section 4.6.

4.2 Theory

As we have explained in the introduction of this thesis (section 1.4.2), PET data can be approximated by a linear model 1.14 (obviating Poisson noise). Neglecting noise and scatter (or considering subtraction from the data), we can simplify the expected activity in a given LOR, $\tilde{y}_i$, as

$$\tilde{y}_i = \sum_{j}^{M} a_{ij} x_j, \tag{4.1}$$

where a_{ij} are the SRM elements, and x_j the image voxel. We have already discussed the physical meaning of the SRM coefficients a_{ij} (see section 1.4.2.1 and 3.2.1), including positron range, non-collinearity, crystal pitch, and any additional effect that makes the non-zero elements lie in a wide region around the LOR, so called tube of response (TOR) [Herraiz et al., 2006a, Markiewicz et al., 2014]. In consequence, the SRM is less sparse than expected. The width of the TOR has a significant impact on image quality, since it limits the best feasible resolution [Moses, 2011] due to the loss of signal-to-noise ratio (SNR) at high frequencies [Herraiz et al., 2007].

In the case of time-of-flight (TOF), it has been proved that longitudinal information about the decay positioning improves the SNR by a factor proportional to the square root of the number of TOF bins within the patient [Conti, 2011]. In this work we propose to subdivide every LOR y_i into L subLORs y_{il} in the transverse direction during the reconstruction process (see figure 4.1), so that

$$\begin{gathered}\tilde{y}_i = \sum_{l}^{L} \sum_{j}^{M} a_{ilj} x_j = \sum_{l}^{L} \tilde{y}_{il} \\ \sum_{l}^{L} a_{ilj} = a_{ij},\end{gathered} \tag{4.2}$$

where a_{ilj} are the normalized elements that define the subLOR y_{il}. We call $\hat{A} \in \mathbb{R}^{NL \times M}$ the SRM matrix for the extended set of subLORs (with matrix elements a_{ilj}).

The rationale of this idea is to ease the inverse problem thanks to the use of narrower TORs, since every subLOR y_{il} is connected with a lower number of voxels. The real challenge here is the estimation of the subLOR number of events, since we do not have direct access to these values (contrary to the TOF data, where the information is obtained experimentally).

Our approach consists on using the expected values of $\tilde{y}_{il}$ as estimators of the relative weight for each subLOR

$$y_{il} = \frac{\tilde{y}_{il}}{\tilde{y}_i} y_i. \tag{4.3}$$

From 4.2 and 4.3 we can easily verify that the total number of events is preserved in each LOR.

It is worth noting that the proposed method is not restricted to subdivisions in the transverse plane. It is feasible to use longitudinal bins or any other approach. The subLORs may also overlap.

The key distinction between subLORs and LORs is that the extended dataset in 4.3 may outperform the resolution limit imposed by the discretization of the measured data in the projection space. The issue is that obtaining a trustworthy activity distribution with low bias and noise is required for estimating the expected values $\tilde{y}_{il}$. In PET, this limit is closely related to the width of the TOR and the sampling of the image space [Moses, 2011].

4.2.1 An algorithm for SR sampling recovery

We can compute the extended data set from the original data by means of an iterative maximum likelihood method [Lage et al., 2015]. In figure 4.2 we show a pseudo code of the method. In the first step, we use the standard MAP-OSEM algorithm to obtain the activity distribution $X^{n_s=0}$. This is the initial image of the super-iterative algorithm to obtain the subLORs using 4.3, and then reconstruct a new image with the appropriate SRM. In this step, each subLOR y_{il} is treated as a new independent LOR for the MAP-OSEM. We call each step in the loop a 'super-iteration'. We have observed in multiple cases that 1 or 2 super-iterations are enough. After that, no further improvement is achieved.

```
$X^{n_s=0} \leftarrow MAP - OSEM_A(\{y_i\})$
for $n_s = 1, N_{super-iterations}$ do
    $\tilde{y}_{il}^{n_s} \leftarrow \sum_j a_{ilj} \cdot x_j^{n_s-1}$
    $y_{il}^{n_s} \leftarrow y_i \cdot \tilde{y}_{il}^{n_s} / \tilde{y}_i^{n_s}$
    $X^{n_s} \leftarrow MAP - OSEM_{\hat{A}}(\{y_{il}^{n_s}\})$
end for
```

Figure 4.2: Pseudo code of the super-iterative algorithm. The matrix A refers to the standard SRM, and $\hat{A}$ refers to the SRM applied to estimate and reconstruct the subLORs. The subindex l indicates the specific subLOR for each LOR i.

We can see this method as a sinogram up-sampling in which the whole data set is used via the reconstructed image. However, instead of explicit interpolations, we distribute the activity inside the TOR preserving the statistics and effectively improving the sampling in the projection domain. The algorithm is very flexible since it deals with each TOR individually. The only assumption is that the SRM of the system is known.

4.3 Simulation study of a simplified model

In this section we show a simple 2D example of the super-iterative reconstruction process implemented in Python. The code can be evaluated from Code Ocean[1]. We have defined a numerical phantom composed of three boxes of 7 pixels size over a circular background, with relative activity 4:1 (subplot a of 4.5 in the results subsection 4.3.1). Four cases were considered: high resolution, low resolution, high resolution with Lagrange polynomials interpolation, and high resolution with the super-iterative algorithm.

1. As a reference of high-quality image, the high resolution sinogram is simulated following three simple steps: image blurring, Radon transform, and Poisson noise addition. These data are reconstructed using the same projection model.
2. As a reference of low-resolution reconstructed image, the high resolution sinogram previously used was down-sampled by averaging every three radial bins, simulating a low-resolution sinogram, i.e. a detector with large crystal size. In this case, the

---

[1] https://doi.org/10.24433/CO.2947710.v2

system resolution is scaled ($\sigma/3$) and the images are up-sampled to the high-resolution image size for comparison.

3. As a reference of well-known method to restore low-resolution data, the low-resolution sinogram has been interpolated using Lagrange polynomials to recover a high-resolution sinogram before reconstruction. The reconstruction scheme for this sinogram is the same as the one used for the high resolution sinogram.

4. Finally, for the super-iterative method we use the image reconstructed from the low-resolution sinogram as the reference image ($X^{n_s=0}$) for the first estimation of the high-resolution extended sinogram ($\{y_{il}\}$). The same reconstruction routine is used to iteratively compute the $\mathrm{n_s}^{\mathrm{th}}$ super-iterated image X^{n_s}, as explained in figure 4.2. The number of super-iterations was set arbitrarily, based on the experimental search of the super-iterations needed to achieve remarkable image improvement.

For standard reconstruction, we have implemented the MAP-MLEM algorithm with optional median root prior and total variation filter for regularization, based on the direct and inverse Radon transform functions of the scikit-image library [van der Walt et al., 2014] and the SciPy library [Virtanen et al., 2020]. The number of radial and angular pixels ($N_\rho \times N_\theta$), the system resolution (σ), the relative activity of the phantom (I), and the reconstruction parameters (number of iterations, number of super-iterations, MAP regularization parameter β, and weight for total variation filter) can be easily modified by the user. The number of image pixels are taken as $N_x = N_y = 6 \cdot N_\rho$.

To quantitatively evaluate the performance of the algorithm, we have computed the mean square error (MSE) and the contrast recovery coefficient (CRC) against the image noise. The MSE is defined in 4.4, where x_i and f_i are the values of pixel i of the reconstructed image and the ground truth image respectively. The CRC (expression 4.5) is given by the average activity in each box of the image ($\bar{X}_{Box}$) divided by the expected activity ($X_{Box} = 4$). The noise (N_{Bg}) is measured as the standard deviation (STD_{Bg}) of the reconstructed activity divided by the mean ($\bar{X}_{Bg}$) inside a given region of the background (Bg).

$$MSE = \frac{\sum_{j=1}^{N} (x_i - f_i)^2}{N}, \tag{4.4}$$

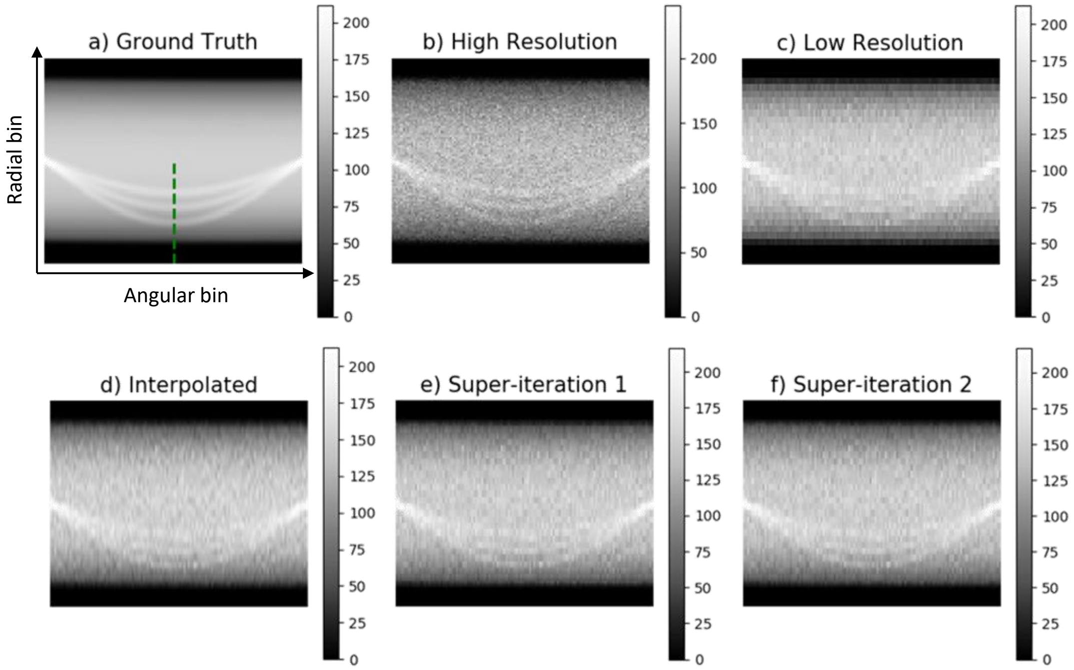


Figure 4.3: Sinograms used to evaluate the super-iterative algorithm with the 2-dimensional example. We show the high resolution sinogram without noise (a) and with noise (b), the low-resolution sinogram (c), low-resolution sinogram interpolated with Lagrange polynomials (d), and the sinograms obtained in the first (e) and second (f) super-iterations of the proposed algorithm. The dashed green line in (a) represents the line profile in figure 4.4.

$$CRC = \frac{\bar{X}_{Box}}{X_{Box}}, \tag{4.5}$$

$$N_{Bg} = \frac{STD_{Bg}}{\bar{X}_{Bg}}. \tag{4.6}$$

4.3.1 Results of the simulation study

For the results shown here, we have set N_ρ=32, N_θ=120, σ=2, I=1, and β=0.005 (no total variation filter was implemented). In figure 4.3 we show the sinograms used for every scenario, and figure 4.4 shows the line profile through the dashed green line. Figures 4.5 and 4.6 show the equivalent reconstructed images and profile through the resolution boxes of the phantom. The image quality improvement can be deduced

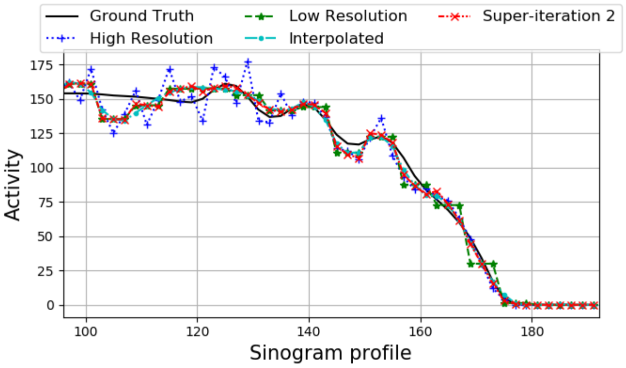


Figure 4.4: Line profile along the dashed green line in subplot *a* of figure 4.3 (super-iteration 1 is not shown).

from the better peak-to-valley ratio in all the boxes, particularly noticeable in the smallest one. The super-iteration 1 has been removed from the profile plots to ease the visualization.

In figures 4.7 and 4.8 we show the MSE and CRC evolution against the noise in a uniform region of the image. We can see the minimum MSE is obtained after a few iterations, while the noise raises continuously with the number of iterations. The CRC reaches a stable maximum for a higher number of iteration. In both cases the MSE-noise and CRC-noise ratios are improved with every super-iteration, but we only show up to the second one since further super-iterations did not achieve a remarkable improvement.

This fact can be understood as the regularized MAP-MLEM iterative algorithm gets stuck after a few iterations in a local maximum of likelihood (while a non-regularized one will end-up producing very noisy images). Super-iterations would work in a similar way to a guided relaxation step in the iterative reconstruction allowing the MLEM to proceed further away from the local solution.

4.4 Materials and methods for PET real data

All the acquisitions presented in this work correspond to the 4R-SuperArgus PET/CT scanner, a former version of the 6R-SuperArgus scanner with 4 detector rings resulting in a shorter axial FOV of 10 cm (instead of 6 rings and 15 cm). We refer the reader to section 2.2.1 for further details of this scanner.

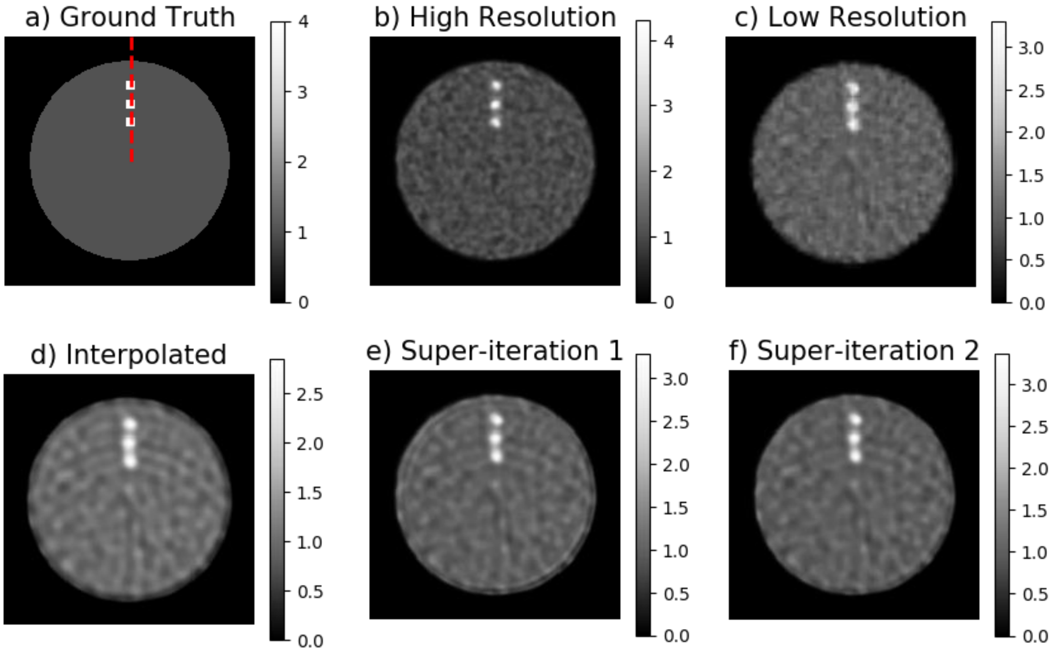


Figure 4.5: 2-dimensional example of the implementation of the super-iterative algorithm. All the images have an equivalent noise level of 0.1. Original phantom (a), high-resolution image (b), low-resolution image (c), image from the high-resolution interpolated sinogram with Lagrange polynomials (d), and first (e) and second (f) super-iterations of the proposed algorithm. The dashed red line in (a) represents the line profile in figure 4.6.

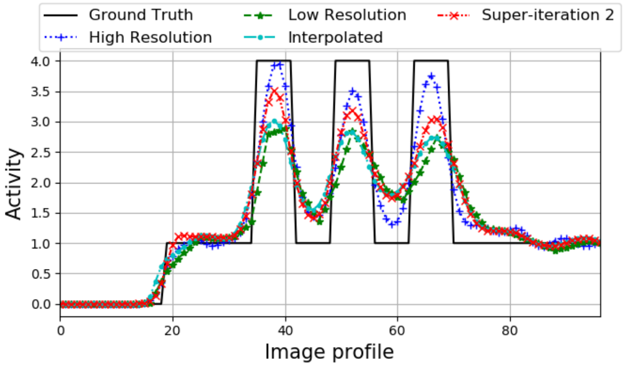


Figure 4.6: Line profile along the dashed red line in subplot *a* of figure 4.5 (profile through all the figures, except the super-iteration 1). The noise level in all the images is equivalent to 0.1.

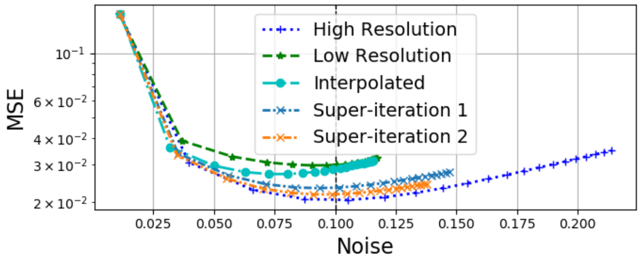


Figure 4.7: MSE of the images in figure 4.5 against the noise per every 10 MAP-MLEM iterations. The iterations increase in the direction of increasing noise. A vertical black-dashed line shows the noise equivalent level of 0.1 taken for the images in figure 4.5 and 4.6.

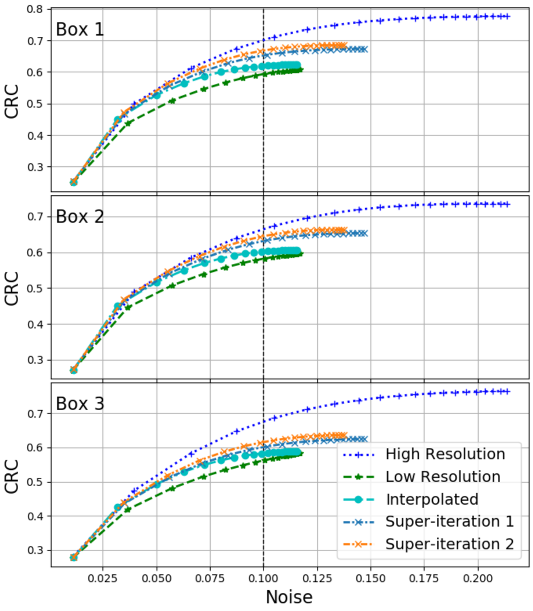


Figure 4.8: CRC of the images in figure 4.5 against the noise per every 10 MAP-MLEM iterations. The iterations increase in the direction of increasing noise. A vertical black-dashed line shows the noise equivalent level of 0.1 taken for the images in figure 4.5 and 4.6.

4.4.1 Acquisitions

First, we quantitatively evaluated the noise and recovery coefficients (RC) using an acquisition with ^{18}F-FDG with the image quality phantom (IQ) from the NEMA NU4-2008 protocol [National Electrical Manufacturers Association, 2008]. The details of the phantom and RC and noise estimation are found in section 2.3. The initial activity was 200 µCi, and the acquisition took 20 minutes, resulting in $7.4 \cdot 10^7$ counts. Attenuation correction was included in this reconstruction.

For comparison, we also show a lower sampling case of the RC-noise curves in which all the counts coming from the GSO crystals (second layer of the phoswich detectors) have been discarded. This acquisition is so called "LYSO layer only". With this approach, higher resolution is achieved at the cost of a higher noise, due to reduced sensitivity. The higher resolution is obtained because the front LYSO layers have less DOI effect than the GSO layers, resulting in a narrower TOR.

The cold Derenzo phantom from section 2.4 is studied here to show qualitative improvement. The initial activity was 400 µCi, and the acquisition took 60 minutes, resulting in $1.8 \cdot 10^9$ counts (notice that the acquisition in the previous chapter 3, section 3.3.2, was performed in the 6R-SuperArgus scanner).

An FDG cardiac mouse heart study is also reconstructed, to show that no artifacts are introduced in in-vivo studies. Acquisitions were done at the Instituto de Investigación Sanitaria Gregorio Marañón (Madrid, Spain). All experimental procedures have been done in compliance with the European Communities Council Directive 2010/63/EU and submitted for approval by the Institutional Animal Care and Use and Ethics Committee of the Hospital General Universitario Gregorio Marañón (HGUGM), supervised by the Comunidad de Madrid according to the Annex X of the RD 53/2013. The subjects were kept at the animal housing facilities of the Unidad de Medicina y Cirugía Experimental (UMCE-HGUGM) in Madrid, Spain. The injected activity was 250 µCi, the acquisition took 20 minutes, and the total number of counts were $4.9 \cdot 10^7$.

The energy window implemented in all the acquisitions was of 425 keV to 600 keV. All the data were acquired in LOR histogram format.

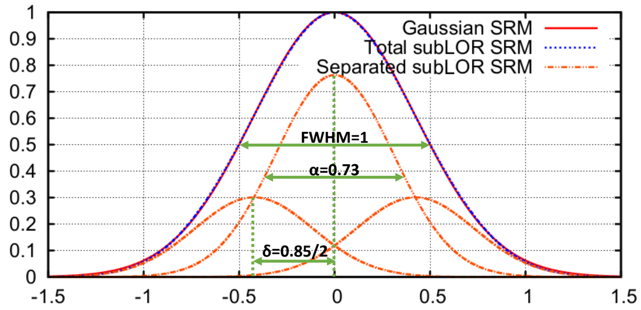


Figure 4.9: Comparison of the Gaussian PSF for the standard reconstruction model (solid red line) with the total PSF for the subLOR reconstruction approach after the fit (dotted blue line). The separated contributions of each subLOR are shown too (dotted-dashed orange line). The parameters obtained are $\alpha = 0.73$ and $\delta = 0.85 \cdot FWHM$.

4.4.2 System Response Matrix

The standard case was evaluated using the GPU-PET reconstructor described in chapter 3, used as a reference, and the SRM detailed in the previous chapter 3. For the extended data set, we modified the FW/BW projection routine to introduce three parallel transverse projections in the radial direction, combined with a narrower PSF (see the scheme in the right-hand side of figure 4.1). If we approximate the PSF image blurring and the projector model to a Gaussian TOR of equal FWHM, this is equivalent to subdivide the TOR into three subTORs (figure 4.9). We have used the following decomposition of the PSF to estimate the distance δ between the subLORs and its relative width α

$$G_\sigma(\rho) \cong N \cdot [G_{\alpha\sigma}(\rho) + q \cdot (G_{\alpha\sigma}(\rho - \delta) + G_{\alpha\sigma}(\rho + \delta))] , \tag{4.7}$$

where G_σ is the Gaussian PSF of width σ, N is a normalization constant, and ρ is the transverse distance from the center of the LOR. We have fixed the weights $q = G_{\alpha\sigma}(\rho = \delta)$. The parameters δ and α describe the SRM of the subLOR extended data set. From a simple fitting experiment, we found that the optimal values were $\alpha = 0.73$ and $\delta = 0.85 \cdot FWHM/2$ (we forced to avoid the trivial solution of $\alpha = 1/3$ and $\delta = 0$). In figure 4.9 we show the resulting approximate TOR and subTORs contribution.

Since the PSF used in the 4R-SuperArgus scanner is an isotropic Gaussian with 1 mm FWHM (0.8 mm FWHM for the LYSO layer only), the three sublors projections are separated a distance of 0.425 mm and the FWHM used in the PSF was of 0.73 mm (0.34 mm distance and 0.584 mm FWHM for the LYSO layer only acquisition).

It is worth noting that neither the SRM in the reference case nor the extended set of subLORs have accounted the asymmetry of TORs due to the DOI. In all the acquisitions under study we are restricted to an internal FOV of 5 cm, where the resolution is more uniform, whereas the total FOV is of 12 cm.

4.4.3 Extended data set calculation

The same projection scheme explained in previous section 4.4.2 is used to estimate the values $\tilde{y}_{il}$ and $\tilde{y}_i$ after each super-iteration. Using the equation 4.3, the extended data set is computed.

4.5 Results

In table 4.1 we show the RC for an equivalent noise level up to the second super-iteration, for both the LYSO-GSO phoswich case and the LYSO layer only case. We refer the reader to section 2.3 to see the details on RC and noise estimation with the NEMA NU4-2008 protocol. We find a RC raise of the 10% for the smaller rod when the super-iteration is applied, whereas the rest of the rods are already closely converged to the unit. The second-superiteration does not show a relevant improvement. For the rods of 4 and 5 mm diameter, we can see a slight overshoot in the single LYSO layer acquisition. In figure 4.10 we can see the improvement in the evolution of the RC-noise per iteration of the first rod.

The reconstruction time for the full LYSO-GSO acquisition shown in table 4.1 took 118 seconds for the standard reconstruction, and 228 seconds for each super-iteration. The FW/BW projection routine took 43 seconds for the standard case, and 153 seconds for the super-iterations (a factor 3.5 extra processing time is required for this routine in the super-iterations). In the LYSO layer only, the FW/BW routines took one fourth of the time, that is the same reduction factor of the number of LORs. In general, the 3.5 extra factor for processing the subLOR projections is preserved for all the reconstructions, although the total reconstruction time may vary (see table 3.1 in previous

LYSO and GSO layers						
4 it. × 9 ss.	1 mm	2 mm	3 mm	4 mm	5 mm	Noise (%)
Standard	0.29	0.90	0.91	0.95	0.96	5.6
Super-iteration 1	0.32	0.92	0.90	0.96	0.98	5.7
Super-iteration 2	0.32	0.93	0.91	0.96	0.98	5.7
LYSO layer only						
4 it. × 9 ss.	1 mm	2 mm	3 mm	4 mm	5 mm	Noise (%)
Standard	0.30	0.89	0.90	1.00	1.00	7.4
Super-iteration 1	0.32	0.89	0.90	1.02	1.02	7.5
Super-iteration 2	0.32	0.89	0.90	1.02	1.02	7.5

Table 4.1: Noise and Recovery Coefficient (RC) for the five rods of the IQ phantom measured following the NEMA NU4-2008 for the Super Argus scanner after 4 iterations of 9 subsets of the MAP-OSEM algorithm (equivalent noise level).

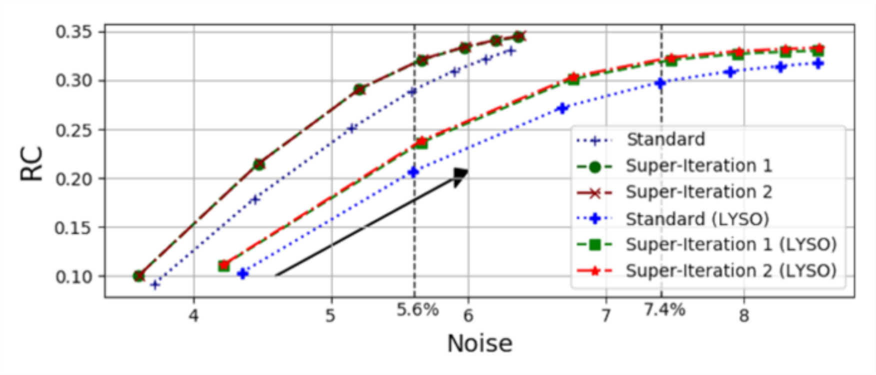


Figure 4.10: RC against noise for the rod of 1 mm diameter of the IQ phantom measured following the NEMA NU4-2008 for the case with phoswich and the single LYSO layer case. Each point represents one iteration of 9 subsets of the MAP-OSEM algorithm, increasing in the direction marked by the arrow. The horizontal dashed line is shown for the noise level of 5.6% and 7.4%.

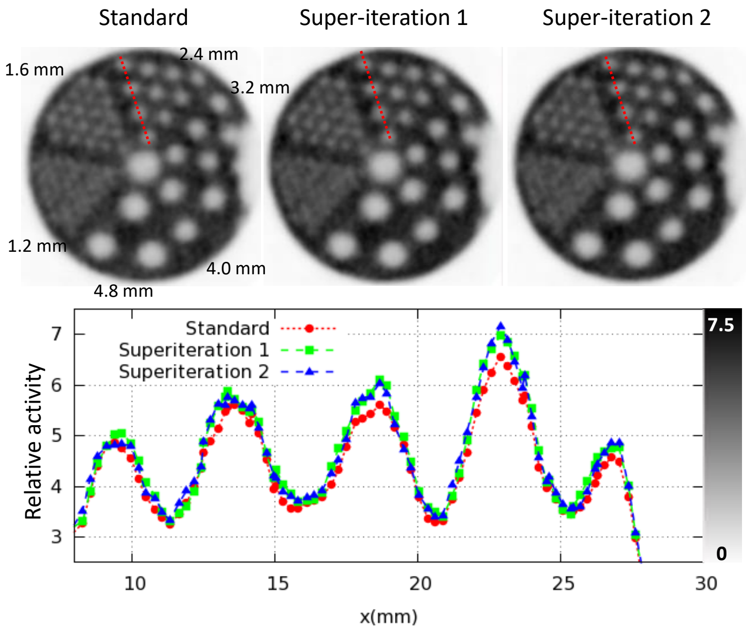


Figure 4.11: Transverse plane of a cold Derenzo phantom for the standard, first and second super-iteration reconstructed for the LYSO layer only case. A line profile along the 2.4 mm diameter rod is shown in the lower panel with the color scale.

chapter 3). The reconstruction software was implemented using PGI CUDA Fortran Compiler in a 24 GB RAM i7 Intel PC with a single 3GB GeForce GTX 780 Ti GPU, with 2880 cores.

In figures 4.11 and 4.12 the reconstructed Derenzo phantom and mouse heart are shown up to two super-iterations. The line profile in the bottom of both images shows the image quality improvement of the resolution for both images.

4.6 Discussion

In this section we have presented a novel method to further improve the image reconstruction in PET. The image space within the FOV is severely oversampled in state-of-the-art scanners with thousands of crystal units resulting in billions of LORs. However, physical detector constraints, such as the crystal pitch, limit the resolution of PET scanners [Cherry et al., 2012, Moses, 2011]. To address this issue, we developed an algorithm that redistributes the coincidences of each LOR into a few subLORs, guided by a previously calculated activity distribution. The maximum likelihood method pro-

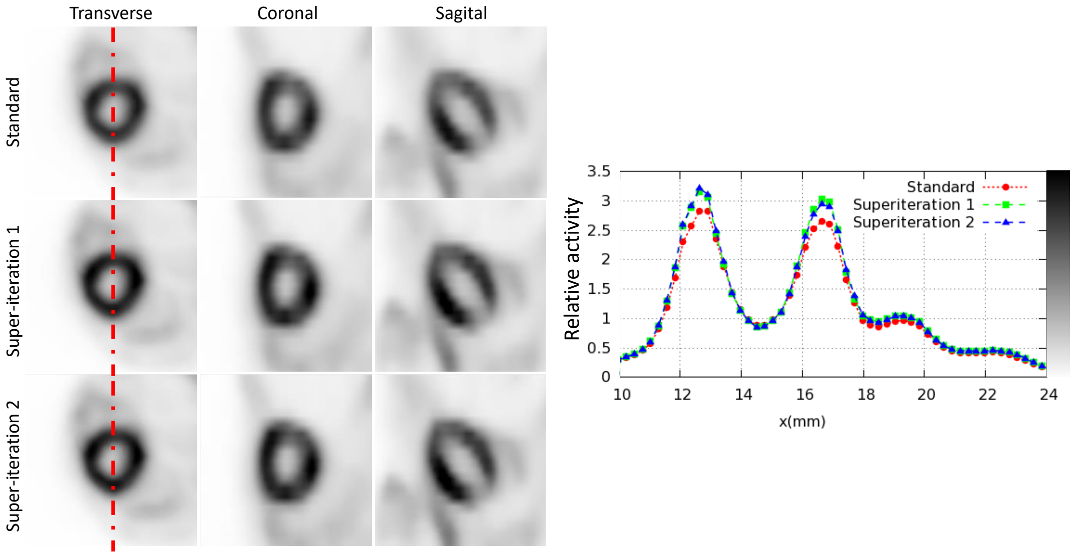


Figure 4.12: Transverse, coronal and sagital planes of a mouse cardiac study 4R-SuperArgus scanner (standard reconstruction, first super-iteration and second super-iteration). The line profile along the dashed line is shown with the color scale (right).

vided by Lage et al. [2015] for the management of triple coincidences is consistent with this redistribution. As a result, we can use the improved data set to rebuild a higher quality image. It is important to remark that this algorithm does not preserve the Poisson statistics. Instead, the subLORs follow a multinomial distribution given by 4.3. Combined with the Poisson distribution of the LOR, we obtain a compound Poisson distribution (CPD), which can be approximated by a scaled Poisson distribution (SPD) [Bohm and Zech, 2014]. Furthermore, the correlation established among the subLORs in 4.3 did not affect the reconstruction negatively, even though we used them as independent variables for the MLEM reconstruction. One has to recall that Poisson statistics is a sufficient condition to derive the actual MLEM equation (see section 1.4.2.2), but it is not a necessary condition. Our algorithm was already successfully applied to other modalities [Galve et al., 2017, Lage et al., 2015, Lee et al., 2018, Moore et al., 2015].

Other authors have studied the benefits of spreading data across sub-crystal units or sinogram sub-bins [Hong et al., 2018, Jin et al., 2012, Kim et al., 2018], but the sampling recovery and count redistribution according to a self-consistent iterative procedure stemming from the current image, based on iterative subTORs in the reconstruction procedure as we have proposed here has never been presented before in PET. The number of subdivisions is not fixed and different schemes can be applied (transverse parallel subLORs were suggested here). X. Jin et al. [Jin et al., 2012] employed normalization-based probabilities to recover axially smashed sinograms, but image based weights with our method could be applied as well. Longitudinal bins combined with TOF information can be implemented too. The method is compatible with multi-ray approaches [Huesman et al., 2000, Moehrs et al., 2008, Zeng et al., 1991] if we consider each ray (or a combination of them) as a subLOR[1]. In this case, the SRM is based on the integration of multiple ray-end points inside the crystals, considering geometrical and physical factors (such as DOI and inter-crystal scatter). Nonetheless, the SRM described in section 4.4.2 keeps the model simple and is accurate enough to demonstrate the viability of the algorithm with just three subLORs.

In SR approaches [Jeong et al., 2011, 2013, Kennedy et al., 2006, Li et al., 2014], previous studies focused on different strategies to improve sampling with small relative displacements between the object and the scanner, like wobbling or stretcher displacements. Although these techniques effectively improve image resolution, they require complex and very accurate motion devices. In contrast, our method does not require any hardware investment that hinders the applicability. Besides, rather than oversampling even more the image space, the algorithm shown improves the discrete sampling, what makes our approach compatible with motion-based SR.

In this work we have demonstrated the algorithm performance for LOR histogram reconstruction, but the flexible method proposed may be easily implemented with any data arrangement. For example, the subTOR divisions are feasible for sinogrammed data. In this case the one-to-one LOR to sinogram bin relation might be lost, because of axial span or transverse rearrangement to match the cylindrical geometry. However, the sinogram bins could be subdivided with the same criteria used in this paper. For list-mode data (in either sinogram format or LOR histogram format), the implementation of the algorithm is straightforward. In this case we can implement a probabilistic strategy based on image projection weights to avoid multiple projections for single coincidence events, which could accelerate the reconstruction. The algorithm 4.2 is compatible with any reconstruction method provided an image to compute the super-iterative weights. We used the MAP-OSEM [Bettinardi et al., 2002, Hudson and Larkin,

---

[1] The subLOR in the sense here defined, since in [Moehrs et al., 2008] the same term was used with a different connotation.

1994, Shepp and Vardi, 1982] since it is a well established standard in PET imaging. It is worth mentioning that the algorithm can also be applied in other emission tomography disciplines such as SPECT, and to any kind of clinical or preclinical PET scanner.

With only one super-iteration, the RC values in table 4.1 increased from 0.29 to 0.32 for the first rod. The GSO crystals provided higher RC-noise ratio compared with the single LYSO layer case, in which they were discarded. The rest of the rods showed RC values close to the unit even for the standard reconstruction. However, after the first super-iteration, we must be cautious to avoid Gibbs artifacts, as was the case with overshoot obtained in rods 4 and 5 of the one-layer acquisition. Overall, we found convergence after the first super-iteration in the real scanner tests. The optimal trade-off between computation time and image quality is achieved by keeping it at just one super-iteration.

The visual quality enhancement of the cold Derenzo after the first and second super-iterations may be seen in the cylinders as an improved peak to valley ratio in figure 4.11. Even though the super-iterations did not diminish the spill over in-between the empty rods, the line profile reveals a better delineation of the walls between the rods. Similarly, after the first super-iteration, the mouse myocardial walls are better defined in figure 4.12. This could lead to better quantification in functional studies or attempting to find tiny lesions.

In theory, a more realistic approximation of the TOR shape would allow to break it into subTORs with no overlap, for a flawless application of the method. However, this requires information of the SRM beyond that predicted by pure geometrical LORs, for example, provided by MC simulations [Herraiz et al., 2006a, Meng et al., 2021]. On the one hand, in section 4.3 we used the same SRM decomposition in the projection modeling and image reconstruction, ensuring the correct TOR definition in our method. The results in this case demonstrated that the method can afford a remarkable image quality improvement when the SRM is optimized. On the other hand, we applied an approximated SRM decomposition based on a Gaussian image blurring kernel and geometrical projections for the realistic scenarios. The original PSF was emulated for the super-iterations by parallel weighted subdivisions of each TOR with narrower blurring kernels (equation 4.7), which was enough to achieve visible image quality gains.

Detailed MC simulations of the emission, transport, and detection of radiation would allow evaluating the proposed method limits. For example, simulations of the asymmetric depth of interaction response for the subTORs can improve resolution in regions close to the edge of the FOV, potentially improving reconstructed images.

We assume that the narrower PSF established for the subTORs contributes to the improved image quality. We know that limiting the number of voxels connected to each LOR increases image quality (TOF-PET vs. non-TOF-PET [Conti, 2011]). In the approach proposed, the distribution of counts in the subTORs is prone to errors due to imperfections in the image employed to estimate the weights (even if the exact SRM was known). This error will propagate to our distribution, compromising the signal-to-noise ratio, and hence limiting the number of subTOR divisions that may be performed. Too many subdivisions may raise the image noise, and too few subTORs may overlap, resulting in minor improvements. The optimal shape and number of subdivisions will depend on the scanner parameters, the activity distribution, and the data statistics, and it should be established experimentally. In some scanners the effect will be more prominent than others, as it was seen from the results in table 4.1. In the 2-D example from section 4.3, the relative improvement was larger for smaller PSF (i. e. less overlap between the subTORs). These results were not published here, but they can be obtained running the code available in the Code Ocean platform.

In terms of processing time, GPUs reduce the reconstruction time by two orders of magnitude [Herraiz et al., 2006a, Zhou and Qi, 2011], allowing the entire process, including super-iterations, to be completed in less than 5 minutes. In any case, the code was not extensively optimized, and the GPU used performs far worse than most modern models.

Chapter 5

Ultra-fast Monte Carlo PET simulator

5.1 Introduction and motivation

Positron Emission Tomography has benefited from Monte Carlo (MC) simulations for decades. Different MC simulation packages have been used for the development and optimization of modern scanners [Aklan et al., 2015, Behnamian and Shafiee, 2018, Delso et al., 2009, Li et al., 2013, Tao et al., 2020]. They have also been used to include realistic physical models in the reconstruction process to improve image quality and help reducing artifacts. For example, the System Response Matrix (SRM) has been approximated with different approaches using MC simulations [Gillam and Rafecas, 2016, Herraiz et al., 2006a, Li et al., 2015, Rafecas et al., 2004, Southekal et al., 2011, Wei and Vaska, 2020, Zhang et al., 2010], and other image corrections such as scatter inside the patient body [Adam et al., 1999, Castiglioni et al., 1999, Ma et al., 2020, Ye et al., 2014], scatter inside the detectors [Abbaszadeh et al., 2018, Lee et al., 2018, Peng et al., 2018], or positron range modeling [Cal-González et al., 2015b, Cal-Gonzalez et al., 2018b, Kraus et al., 2012] have been addressed. Particle therapy has also paid attention in PET (and other imaging techniques, such as Prompt Gamma detection) for non-invasive dose monitoring. Nuclear activation during irradiation generates positron emitter fragments that can be potentially used for range verification [Bauert et al., 2019, Kraan, 2015, Marcatili et al., 2020, Masuda et al., 2019, 2020, Zhu and Fakhri, 2013]. MC simulations of PET signals are required to have a reliable estimation of the detector response and the image reconstruction in clinical scenarios [Augusto et al., 2018, Choi et al., 2020, Jan et al., 2013, Onecha et al., 2020, Valladolid Onecha et al., 2021].

Since the first known reference on Monte Carlo methods in the letter from J. von Neumann to R. Richtmeyer in 1947 about the method for neutron diffusion [Ulam et al., 1947], many different MC based codes developed for particle tracking simulations can be found in the literature. In recent times, some of the most popular ones are the Electron Gamma Shower (EGS) code system [Kawrakow, 2000a,b, Rogers, 1984] or PENELOPE [Agency, 2019, Baró et al., 1995], both for coupled transport of electron and photons, GEANT4 [Agostinelli et al., 2003, Allison et al., 2016], a simulation toolkit for the passage of many particle types (protons, neutrons, electrons, etc.), the Monte Carlo N-Particle (MCNP) code [Briesmeister, 2000, Werner et al., 2017] for neutron, photon, and electron transport, and FLUKA [Böhlen et al., 2014, Ferrari et al., 2005] for interaction and propagation in matter of about 60 different particle types.

These simulation platforms have been widely used beyond medical physics, and specific toolkits have been developed to simplify their applicability in this field. One of the most common open source software developed for medical imaging and radiotherapy is GATE [Jan et al., 2011, 2004, Sarrut et al., 2021, 2014], which is based on GEANT4. SimSET [Baum and Helguera, 2007, Guerin and El Fakhri, 2008, Poon et al., 2015] (an acronym for Simulation System for Emission Tomography), is a Monte Carlo simulation package for emission tomography (PET and SPECT) based on variance reduction tools to enhance computational efficiency. Due to its high performance with voxelized geometries, SimSET was also combined with other simulation packages in the past, like MCNP [Du et al., 2002, Mok et al., 2010], GEANT4 [Barret et al., 2005], or GATE [Chen et al., 2008, Lin et al., 2014]. Recently, SimSET has been integrated in a user-friendly platform called SimPET [Paredes-Pacheco et al., 2021] which provides a graphical user interface to ease the access to the simulations. An easy-to-use adaptation of PENELOPE to PET systems has been implemented in PeneloPET [Abushab et al., 2016, Cal-González et al., 2013, España et al., 2009, Lopez-Montes et al., 2019]. PeneloPET is a simple to configure code for many different scanners since it works with a few simple input text files. However, it is more cumbersome to use in the case of complex geometries that are not based on the classical cylindrical shape, as in relies on PENELOPE. PETSIM [Thompson et al., 1992], Eidolon [Zaidi et al., 1998], or Gray [Freese et al., 2018] are other examples of MC PET simulators. Further information about the development of MC codes dedicated to medical physics, and dedicated PET-SPECT codes can be found in [Buvat and Lazaro, 2006, Rogers, 2006].

There are also other approaches beyond MC methods, based on pseudo analytical simulators [Comtat et al., 2003]. Detailed MC calculations are computationally expensive, and the simulations may be impractical for some applications. For this reason, pseudo analytical simulators use simplified models to accelerate the performance (analytical projections together with statistical methods to implement noise, random coincidences, or activity decay, detector efficiencies, attenuation, spatial resolution, and added scatter background). Among the most recent ones, we can mention STIR [Thielemans et al., 2012], PETSTEP [Berthon et al., 2015], or SMART [Pfaehler et al., 2018].

Even though pseudo-analytical methods might be very practical for certain applications, MC methods are still the gold standard in terms of accuracy. An effective solution to overcome long execution times without applying simplifications is the use of general purpose graphical processing units (GPUs) which allow parallel computing in thousands of thread processors, thus increasing the overall code efficiency at the expense of higher programming effort. There are many examples of the implementation of GPU parallelization in MC codes for different particle tracing scenarios that have been released along the last fifteen years. Among the literature we can find independent Monte Carlo codes like [Alerstam et al., 2008b], based on the White Monte Carlo developed in [Alerstam et al., 2008a], the code from [Badal and Badano, 2009] for photon tracking in the energies between 50 eV to 1GeB, the GPU Monte Carlo dose code for coupled photon-electron transport in the range 0.01-20 MeV [Hissoiny et al., 2011], or gPMC for proton dose calculation [Jia et al., 2012a]. Other authors relied in previously developed CPU-based MC simulation codes, like the GPU implementation of EGSnrc [Lippuner and Elbakri, 2011], the GPU versions of DPM (Dose Planning Method) [Sempau et al., 2000] so called gDPM [Chi et al., 2016, Jia et al., 2011, 2010, 2012b], or different versions derived from Geant4 [Bert et al., 2013, 2012, Garcia et al., 2016, Jahnke et al., 2012]. However, if we talk about PET dedicated codes, these are much fewer. To the best knowledge of the author, we just found three packages for total PET simulation. The first one is based on the accelerated version of Geant4, GPU accelerated Geant4 based Monte Carlo Simulation (GGEMS) [Ma et al., 2020]. The second one is gPET [Lai et al., 2019], based on gDPM. Both perform the entire process since the positron emission to the photon detection using GPU parallelization, even though [Ma et al., 2020] used a hybrid CPU-GPU method for the generation of single events and coincidence sorting. The last one is MCGPU-PET[1][Lopez-Montes, 2021], based on the previous work from [Badal and Badano, 2009]. The main flaw of MGGPU-PET is the missing particle transport through the detectors and detector response model.

One of the main goals of this thesis was to develop a extremely flexible, multiple-purpose fast GPU based Monte Carlo PET simulator. We were looking for a flexible tool that can be used to simulate any scanner geometry and assist in the image reconstruction in different approaches, such as scatter correction, improving the system response matrix, or direct implementation of the simulator in the projection step of the reconstruction process. The simulator should also be able to accurately estimate certain fundamental quantities for the performance of the scanner, such as the spatial resolution and the sensitivity, in order to extend its application to the scanner design. We developed the Ultra-fast Monte Carlo PET simulator (UMC-PET) [Galve et al., 2020a] to attain these goals while optimizing speed, flexibility and accuracy in a code as simple and modular as possible.

In this chapter, the UMC-PET code is explained in detail, and several validations and benchmarks are presented. In section 5.2.1 the input files regarding scanner and source description and general simulation parameters are introduced. We explain the code details, within its main routines, in section 5.2.2. In section 5.2.3 we explain the general output files. Validations against PeneloPET are presented in section 5.3 and an example of its application is presented in section 5.4. Finally in section 5.5 we review the performance of the simulator and discuss future improvements.

5.2 Description of the UMC-PET simulator

The overall workflow of the UMC-PET can be seen in figure 5.1. The code starts reading the general parameters of the simulation, the scanner and objects images, and the source image. The materials cross sections have been obtained from PENELOPE [Agency, 2019], as well as the dispersion angles and energies for photon scattering. One of the key characteristics of the code is the use of a voxelized definition for both the scanner and the object, giving full flexibility to define scanner geometry and detector shape. We used the NVIDIA CUDA platform and PGI CUDA Fortran Compiler for the GPU

---

[1] https://github.com/DIDSR/MCGPU-PET

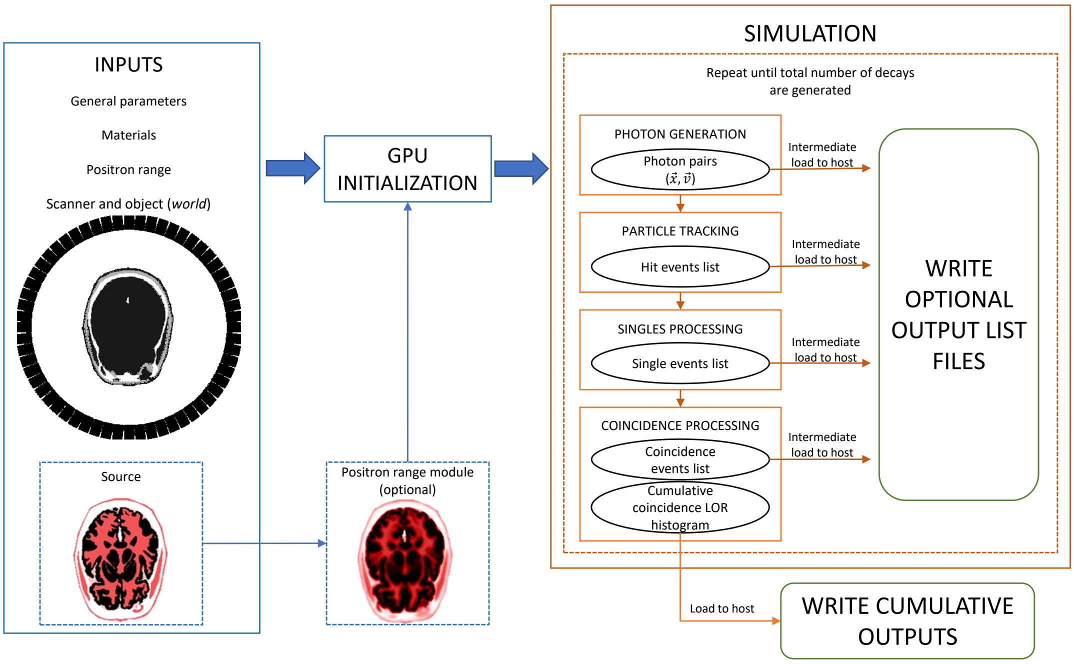


Figure 5.1: Schematic representation of the workflow of the UMC-PET simulator.

implementation, defining different CUDA kernels for particle generation, photon tracking, singles processing, and coincidence processing. In the first routine, two photons are initialized per thread from a random annihilation point. In the second one, each thread of the GPU will follow the photons through the space until they leave the simulation space or they are totally absorbed, generating a list of hit events that is passed to the singles generator. Finally, the singles are paired in the coincidence processing routine. This implementation reduces the complexity of the coincidence shorting, at the expense of missing timing evolution and discarding random coincidence or pile-up events. The code offers cumulative outputs (lines of response (LOR) histograms for separate true and scatter events), and intermediate outputs, such as list of hits, list of singles, or list of coincidences can be defined.

5.2.1 Input files

The development of the code was oriented to a compact and flexible description of any scanner geometry, preventing modifications of the code from one scanner simulation to another. In this sense, we opted for a definition of the scanner that does not imply any assumption of its geometry or its electronics, besides keeping in computer memory as much information as possible in order to avoid computation of any index inside the code. Generating the scanner files implies additional work from the user point of view using independent tools, and additional software based on the PeneloPET input files was prepared to ease simple scanner simulations. The input files defining the simulation might be separated into scanner and objects files, source file, materials file, and general parameters file.

5.2.1.1 Scanner and object definition

- Scanner and object image

  We chose a voxelized definition of the *world* inside the simulator. As shown in fig. 5.2, the scanner detectors and any additional object (like shielding, the patient table, or the patient body) are included in an image that is read by the code. The image is allocated in the computer, given the array dimensions and physical dimensions stated by the user. In the image, the scanner blocks are numerated with positive numbers, whereas other objects (such as patient body, shielding, etc.) are defined with negative indices referring to each material.

- Detector image

  We have separated the image in scanner blocks and the crystals inside a block. The blocks are also voxelized in two high resolution images of equal dimensions that define the crystal index inside the block and the crystal material. This definition allows us to accurately delimit the traditional rectangular prism-shaped blocks (like the SuperArgus [Wang et al., 2006] or the Biograph mMR detector unit [Delso et al., 2011]) due to the symmetry with the image voxelization (i.e. one pixel per crystal), discarding any overlapping error. Furthermore, complex crystal distributions like the hexagonal pixel shown in [Perez-Benito et al., 2018] can be easily implemented in the code.

- Detector coordinates

  The position and orientation of each block is read from a different file, giving the central coordinates of one of the faces of the block and three vectors that define the orientation of the block. In fig. 5.3, the position vector is referred as $\vec{R}_{block}$,

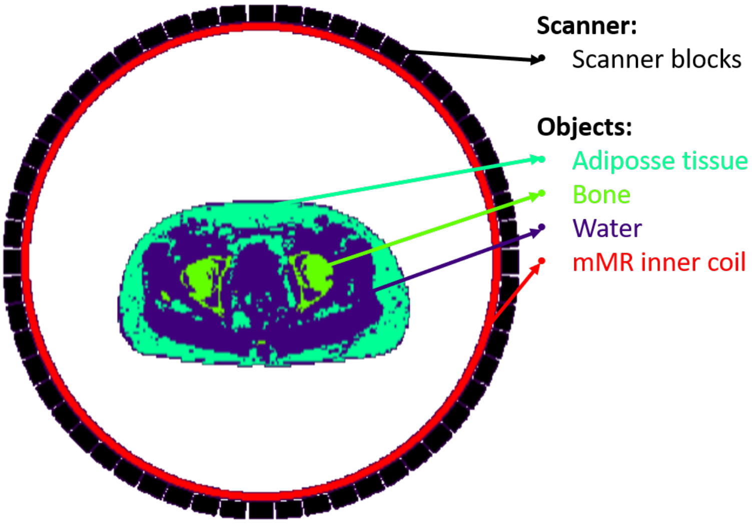


Figure 5.2: Scheme of the mMR scanner with the inner coil used for magnetic resonance inside the PET bore and a pelvis CT segmented in three materials: adipose tissue, bone, and water. The blocks are numbered with positive indices, whereas the coil and the CT materials are defined with negative numbers.

while the orientation vectors are referred as $\vec{v}_x$ and $\vec{v}_y$ (the third vector $\vec{v}_z$ is out of the 2D image). These vectors must form an orthonormal basis in the cartesian space centered in the position vector, and they must be oriented so that a simple projection of the vector $\vec{R}_{hit}$ over each one of them gives the position of the hit inside the block image.

- Coincidence matrix

  The coincidence matrix specifies the couples of detectors that are found in coincidence. The matrix size is equal to $N_{detectors} \times N_{detectors}$ (where $N_{detectors}$ is the total number of detectors in the scanner), and the value of each element is the LOR index for that couple. The pairs of detectors that are not in coincidence are marked with a negative index. Since most of the scanners are cylindrical, the code also allows to define separately the number of rings, and the number of blocks in a ring, in order to organize the matrix per each couple of detectors.

  In fig. 5.4 we can see an example of the coincidence matrix we will use later to simulate the 6R-SuperArgus scanner [Udias et al., 2018] and a prototype of the HSTR-BrainPET [Catana, 2019a,b, Galve et al., 2020b]. Only the upper

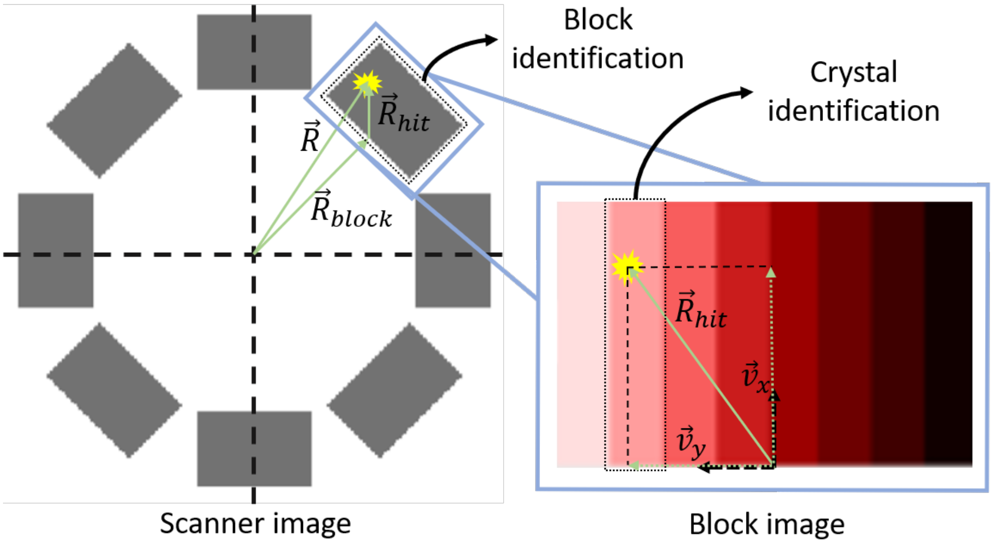


Figure 5.3: Scheme of the crystal identification after a hit occurs when the block is separately pixelated of the scanner image.

diagonal of the combinations inside a ring is considered, since the code transposes the detector indices when they are below the diagonal. In the 6R-SuperArgus scanner, which has cylindrical geometry, the matrix is separated into 6×6 equal submatrices of 24×24 detectors corresponding to the combinations for each couple of rings. The HSTR-BrainPET has a spherical geometry, and its matrix was limited by the angle between each couple of detectors. This fact makes the matrix much more complex.

- Look-up tables and Anger logic

  The code reads a file with the centroid of each crystal in the blocks image, and a 2D image that represents the look-up table for the crystals. The centroids are written in pixel coordinates inside the look-up table image, so that we can apply the Anger logic to the XY coordinates of multiple-hit singles and directly obtain the crystal designated (see section 5.2.2.5). Again, this framework will be very useful when the detectors are not the traditional prism-like detector. There is no XY blurring incorporated in the signal obtained by the detector readout system.

- Multi-layered detectors

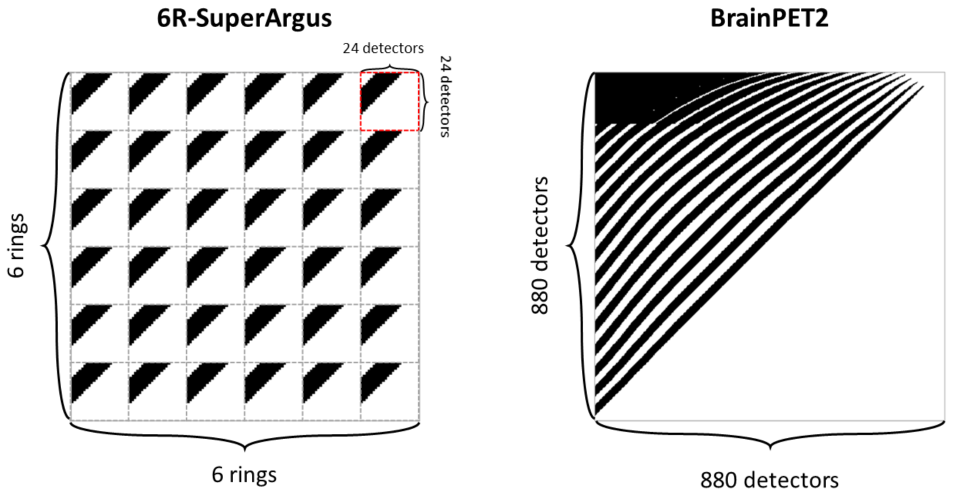


Figure 5.4: Coincidence matrix representation of the 6R-SuperArgus scanner and the HSTR-BrainPET scanner. The detectors pairs in coincidence are marked in black.

Although it is possible to identify multi-layered crystal arrays using only the Anger logic and the appropriate look-up table [Mohammadi et al., 2017], there are detectors where a few crystals of different layers share the same centroid (like the SuperArgus module [Udias et al., 2018, Wang et al., 2006]). In these cases, we may define the number of layers to separate the singles at different depth points in the detector by measuring the energy deposition at each layer.

- Energy resolution

  The energy resolution is separately defined for each crystal inside a block, so that we can have crystals with different characteristics inside a detector, like the phoswich scintillators used in the SuperArgus scanner [Udias et al., 2018, Wang et al., 2006].

5.2.1.2 Source definition

The source definition also uses a voxelized image. The memory size of the array to be allocated, its physical dimensions, and the center position of the source are read as input. Each voxel of the image represents its relative activity (no units used), so the array represents the activity distribution in the space to randomly select the emission position for each decay.

It is possible to choose a radionuclide in the source parameters that will be used to select the positron range (PR) blurring kernel implemented to the source activity distribution. The PR kernel was first introduced by [Cal-González et al., 2015b], and it is explained in more detail in section 5.2.2.2.

5.2.1.3 Number of decays

UMC-PET does not simulate a given activity and decay time, but a number of decays selected by the user N_{decays}. However, the number of positron annihilations to be simulated is easily scalable for a given radionuclide, initial activity, and acquisition time if we apply the exponential decay rule and the positron emission branching ratio of the radionuclide. Given an acquisition time T, initial activity A_0, decay constant λ, and β^+ branching ratio Γ_{e^+}, the number of decays N_{decays} reads as follows:

$$N_{decays} = \Gamma_{e^+} \cdot \int_0^T A_0 \cdot e^{-\lambda t} dt = \Gamma_{e^+} \frac{A_0}{\lambda} \left(1 - e^{-\lambda T}\right). \tag{5.1}$$

We used this formula in the cases we wanted to have a realistic noise level in the data, but we might include a database of common PET radionuclides in the future to automatically set the number of particles after an input acquisition time, initial activity, and radionuclide selection.

5.2.1.4 Materials

In section 5.2.1.1, we already mentioned the material definition of the objects inside the scanner. We are interested in the photon mass attenuation coefficients of these materials to estimate the probability of an interaction to occur, and specifically the probability of each kind of interaction for high energy gamma photons: coherent scatter (Rayleigh scatter), incoherent scatter (Compton scatter), photoelectric absorption and pair production. In practice, we are going to work with photons up to 511 keV, so the pair production will not be accounted (due to energy conservation, we need photons of at least 1022 keV to have a pair production event).

PENELOPE (PENetration and EnergyLOss of Positrons and Electrons) [Agency, 2019, Baró et al., 1995] is a Monte Carlo simulation code for coupled electron-photon transport within an energy range from 50 eV to 1 GeV. It is an open source code developed by the Universitat Politèctica de Catalunya (UPC) and distributed by the Nuclear Energy Agency (NEA). We use PENELOPE to obtain all the required coefficients for our UMC-PET simulator. The auxiliary program *MATERIAL* is employed to extract atomic interaction data from the database of PENELOPE, and the program *tables.f* is used to access the mass attenuation coefficients required. These coefficients have to be interpolated between the energy range of interest, from 1 keV to 1 MeV, which will be read by the code. In figure 1.3 (introduction section 1.1.2), the coefficients for water and LYSO from 1 keV up to 100 MeV are shown.

We also use PENELOPE to calculate the deviation angle and energy deposited after each scatter event. To reduce calculations, *penmain* is used to precompute a sample of events per energy bin of 1 keV, from 1 keV to 1MeV. This sample of events is read by the code, and it will be used later to randomly select the final state of a particle after a scatter event occurs.

5.2.1.5 GPU control parameters

We have defined input parameters to control other features about GPU performance, like the number of threads ($n_{threads}$) and blocks (n_{blocks}) that we use in every call to the GPU, the number of decays generated per thread (n_{decays}), or the number of hits/singles kept in the GPU memory per particle (n_{hits}). Since the GPU multiprocessor groups threads in warps of 32 threads [Fatica and Ruetsch, 2014], we strongly encourage to use a multiple of 32 for the number of threads. The optimal number of blocks and threads to be used will depend on GPU model and simulation images (both the scanner and object image, and the source map), so we recommend performing a few trials before running long simulations. In our tests, the optimal number of decays per thread was simply one ($n_{decays} = 1$), but the parameter remains available for performance experimentation. The number of hits/singles per particle depends on the simulation, but very seldom more than ten interactions per particle occur.

5.2.2 Description of the code

In this section we describe the main routines involved in the PET simulation scheme presented in figure 5.1 and other main routines like the Random Number Generator. In the simulation step, the routines for photon initialization (5.2.2.3), photon tracking

(5.2.2.4), singles generation (5.2.2.5), and coincidence sorting (5.2.2.6) work sequentially, forwarding the output of each routine to the input in the next one. Each subroutine works with a batch of decays defined by the number of threads and blocks times the number of decays per thread ($N_{batch} = n_{decays} \times n_{threads} \times n_{blocks}$, see section 5.2.1.5). The number of calls is given by the number of decays (5.2.1.3) divided by the decays batch size in a thread ($N_{calls} = N_{decays}/N_{batch}$).

5.2.2.1 Random Number Generator

The Fortran implementation of the subroutine RANECU in [James, 1990] is used for pseudo-random number generation. L'Ecuyer [L'Ecuyer, 1988] previously introduced the algorithm, proving a uniform distribution of pseudo-random numbers in the range $[0, 1)$ with a period of $\approx 2 \cdot 10^{18}$ is obtained. To guarantee the independence of the distributions among different threads, millions of seeds were precomputed, and transferred to each thread on the GPU.

5.2.2.2 Positron Range

The Positron Range (PR) has a direct impact in PET image quality [Alva-Sánchez et al., 2016, Cal-Gonzalez et al., 2009, Park et al., 2007, Peng and Levin, 2012] which cannot be neglected in realistic simulations. An accurate representation of the PR requires the simulation of the radionuclide decay chain, the positron energy spectrum, and the positron interactions and track through the medium, that is going to enlarge the computational burden. In the UMC-PET we have implemented a simplified PR modeling using the analytical expression presented in [Cal-Gonzalez, 2014, Cal-González et al., 2013] to estimate a spatially variant PR kernel based on the density map, similar to the approach successfully applied in [Cal-González et al., 2015b, Cal-Gonzalez et al., 2018b] for PR correction during image reconstruction. The authors showed good agreement against reference simulations when the annihilation image is homogeneous, although small deviations could be found in the interfaces between materials. The PR kernel is applied to the activity image before the simulation (see figure 5.1), to generate an annihilation map that is loaded in the GPU global memory for the simulation step.

In Cal-Gonzalez [2014], Cal-González et al. [2013], the authors presented several representations to model the PR. The simplest one is the annihilation point spread function (aPSF), which is defined as the annihilation probability at a given point of space. From the aPSF we can define the 3D radial histograms,

$$g_{3D}(r) \equiv 4\pi \int_{r}^{r+dr} r'^2 aPSF(r')dr'. \tag{5.2}$$

And the authors defined the following analytical expression for the $g_{3D}(r)$:

$$g_{3D}(r) \approx C\left[(a\cdot r+1)\left(1-\frac{r}{r_0}\right)^n - \frac{\epsilon}{r^n}\right], \tag{5.3}$$

where r_0 is the maximum PR radial distance ($g_{3D}(r) = 0\ if\ r > r_0$). The rest of the parameters are fitting parameters with no specific physical meaning. We can differentiate two terms in the equation, governing the limit of $r \to 0$ and $r \to \infty$ separately. This representation gives flexibility enough to accurately fit the parameters to the simulated distributions.

In table 5.1 we show the values reported in [Cal-González et al., 2015b] of the parameters a, r_0, n, and r_i (given that $\epsilon \equiv 0.01 \times r_i^n$) for different radionuclides in water. The authors obtained these values fitting equation 5.3 to simulations performed with PeneloPET (see [Cal-Gonzalez, 2014, Cal-González et al., 2013] for further fitting details). The constant factor C is a normalization factor which can be calculated prior to the convolution. In figure 5.5 we show a representation of equation 5.3 using the values given in the table 5.1.

In Cal-González et al. [2013], the authors proved that equation 5.3 can be scaled to any other tissue based on its relative electronic density, using the variable change $r_e = r\cdot\rho/\rho_{water}$. In Cal-González et al. [2015b], a spatially variant kernel was proposed using the density map from a CT to adapt the aPSF for non-homogeneous media. We use a similar approach, but we use $\mathrm{g_{3D}}$ instead:

$$\begin{aligned}\tilde{x}\left(\vec{r}\right) &= \left(x * g_{3D}\right)\left(\vec{r}\right) = \int_{\theta,\phi} d\theta d\phi \int_{r_e} x\left(\vec{r}+\vec{r}_e\right) g_{3D}\left(r_e\right) dr_e \\ C^{-1} &= \int_{r_e} g_{3D}\left(r_e\right) dr_e,\end{aligned} \tag{5.4}$$

where x is the original activity map, $\tilde{x}$ is the blurred activity from the convolution with $\mathrm{g_{3D}}$, and C is the normalization constant. The variable r_e is integrated in the direction given by (θ, ϕ), accounting for the local density and thus it is locally dependent for each point $\vec{r}$ and each integration angle (θ, ϕ). To discretize the continuous integral,

Radionuclide	$a(mm^{-1})$	$r_0(mm)$	n	$r_i(mm)$
^{18}F	2.60	2.39	4.04	0.051
^{11}C	1.48	3.80	3.16	0.21
^{13}N	1.46	4.95	3.04	0.29
^{15}O	1.15	7.90	3.09	0.36
^{68}Ga	2.41	8.98	3.27	0.20
^{124}I	3.07	11.0	4.46	0.36
^{94m}Tc	2.49	11.1	2.72	0.33
^{82}Rb	7.13	16.6	3.11	0.41

Table 5.1: Parameters of the fit to the g_{3D} distribution of PeneloPET to equation 5.3 taken from [Cal-González et al., 2015b]. r_0 is the maximum PR radial distance ($g_{3D}(r) = 0 \ if \ r > r_0$). The rest of the parameters are fitting parameters with no specific physical meaning.

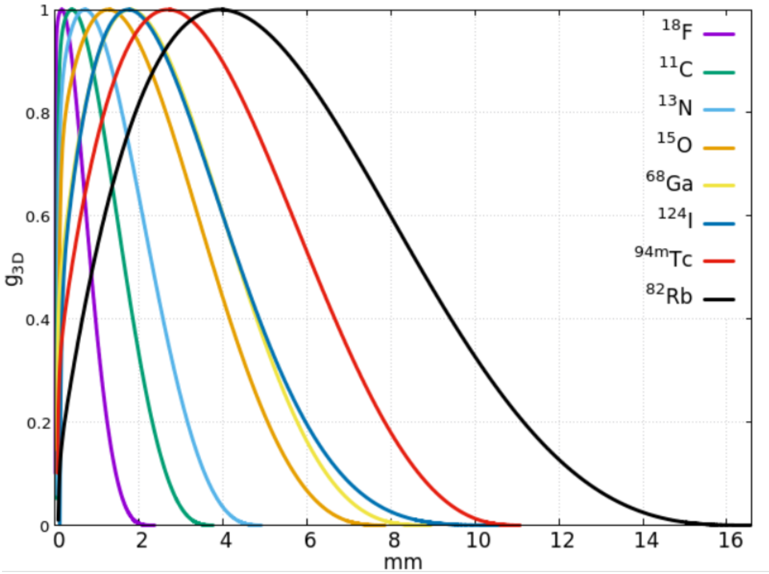


Figure 5.5: Plot of the analytical expression for the g_{3D} of the PR given by equation 5.3 using the parameters of the table 5.1.

we perform a line integral increasing r_e a variant step size $\Delta r \cdot \rho_j/\rho_{water}$ (where ρ_j is the density at a voxel j) until $r_e > r_0$. The normalization factor C is independently estimated for each angular direction (θ, ϕ) since the path through different materials may result in numerical variations of the total integral. A free parameter is left to select the number of steps per pixel (N_{st}) by the user that defines the step size in the line integral ($\Delta r = min(s_x, s_y, s_z)/N_{st}$, where $s_{x,y,z}$ refers to the pixel size in each coordinate direction). The number of integration angles N_θ and N_ϕ are chosen so that we get an average of N_{st} integration points per pixel at the sphere surface of radius r_0 ($N_\theta = \pi \cdot r_0/min(s_x, s_y, s_z) \cdot N_{st}$, $N_\phi = 2\pi \cdot r_0/min(s_x, s_y, s_z) \cdot N_{st}$). The convolution was implemented in the GPU, using a thread per each voxel in the image.

In fig. 5.6 we show an example of the blurring kernel in heterogeneous phantom of bone, air and water, for two different radionuclides, ^{18}F (2.39 mm maximum radius in water) and ^{82}Rb (16.6 mm maximum radius in water).

5.2.2.3 Particle Initialization

In the first simulation step of UMC-PET, all the threads are initialized with a batch of antiparallel photons. The PR blurred image from previous section 5.2.2.2 is taken to randomly pick an emission voxel using the Walker's aliasing method (see section 1.5.2.5) and the physical emission point is randomly selected with a uniform distribution inside the voxel volume. The direction of emission (ϕ, θ) of the first photon is randomly selected with a uniform distribution in $\phi \in [0, 2\pi)$ and $\cos\theta \in [-1, 1)$. The second photon is selected with opposite direction, and rotated an angle taken from a Gaussian distribution of 0.5 degrees FWHM to simulate non-collinearity (section 1.1.1.2). At the end of this routine, each thread saves in the global memory of the GPU the annihilation point and the direction vectors for both photons for every decay in the thread.

5.2.2.4 Photon Tracker

In this routine the two photons for every annihilation are tracked in a single thread until they are totally absorbed or they are out of the scanner. All the interactions are saved in the global memory of the GPU device (up to the maximum number of hits defined in the control parameters, section 5.2.1.5) that is read by the Singles Generator routine (section 5.2.2.5). The material index (negative for objects, positive for the scanner detectors), the energy deposited, and the time of flight (TOF) since the annihilation

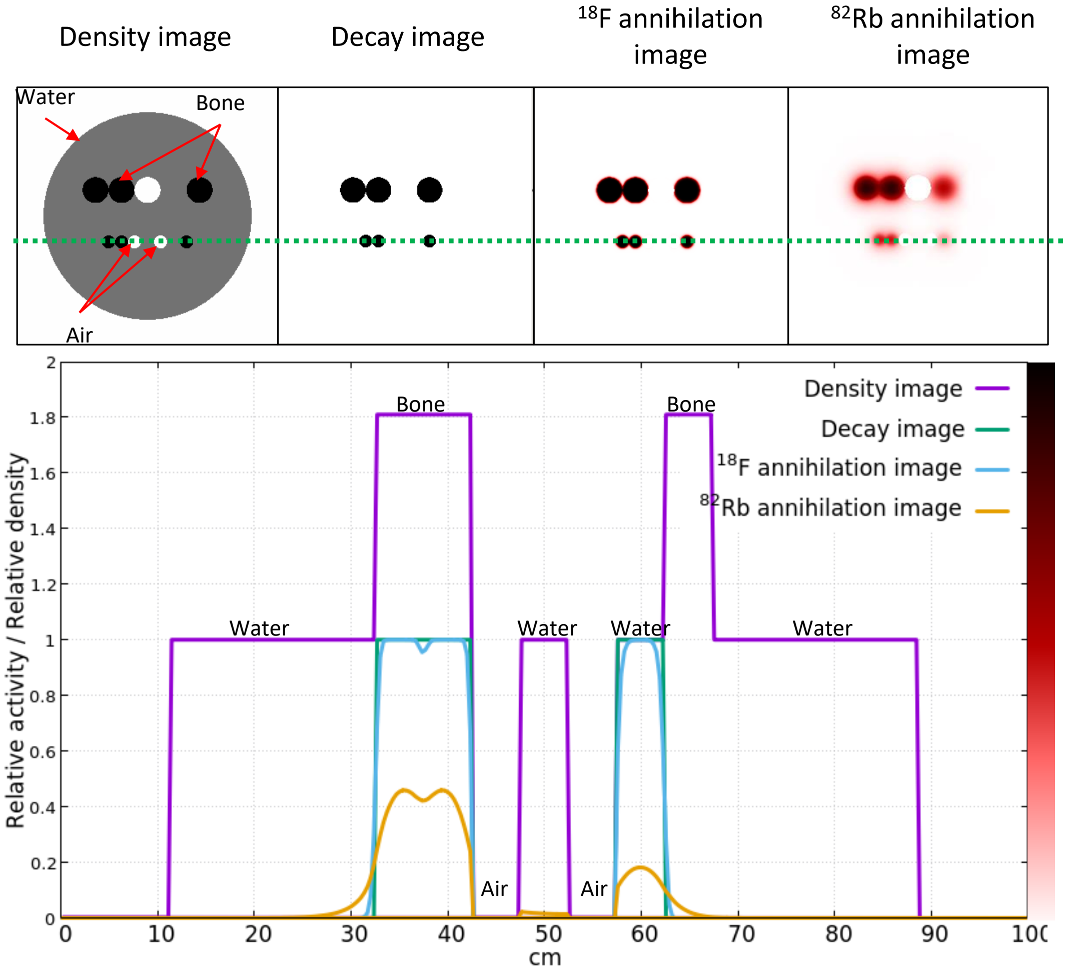


Figure 5.6: Effect of the positron range blurring kernel in a simulated phantom with different sphere sizes (1 cm and 0.5 cm diameter) and materials (water, air and bone). We show the image of relative densities (top left), the original decay image (top middle-left), and the blurred images using ^{18}F (top middle-right) and ^{82}Rb (top right). The line profile along the green dotted-line is shown at the bottom of the figure.

event are saved. Additional information such as the emission voxel or the hit voxel are also saved to have more knowledge about the simulation.

The Woodcock algorithm [Badal and Badano, 2009, Carter et al., 1972, Hissoiny et al., 2011, Woodcock et al., 1965] for particle tracking has been implemented. This algorithm considers the particle is always inside a reference material with attenuation coefficient μ_{ref}, that is a combination of the actual material $\mu\left(\vec{x}\right)$ and a fictitious material so that $\mu_{ref} = \mu\left(\vec{x}\right) + \mu_{fictitious}$ (notice implicit energy dependence of the attenuation coefficients). At every step, the particle advances a distance d derived from μ_{ref},

$$d = -\frac{ln\left(\xi\right)}{\mu_{ref}}, \tag{5.5}$$

where ξ is a uniform random number distributed in the range $[0, 1)$. The reference material is selected by the user, and it should match the most dense material in the simulation. Note that $1/\mu$ is the mean free path for a given material.

To determine whether an interaction occurred, we compute a second random number, and the probability to have an interaction given by the reference mean free path and the attenuation coefficients,

$$P_{hit} = \mu\left(\vec{x}\right)/\mu_{ref}. \tag{5.6}$$

We use a sequential search (section 1.5.2.4) to determine which kind of interaction takes place, where the probability is given by its attenuation coefficient and equation 5.6, and the total attenuation is the sum of each contribution. We have four different interactions:

- Fictitious interaction. There is no change in the energy and direction of the particle.
- Photoelectric effect. The photon is absorbed.
- Compton interaction. The photon changes its energy and direction. Both deposited energy and the dispersion angle θ are selected among precomputed events, and a ϕ angle uniformly in the range $[0, 2\pi)$.
- Rayleigh interaction. The photon conserves its energy, but it changes its direction. The new θ angle is randomly selected among precomputed events at each energy, and a ϕ angle uniformly in the range $[0, 2\pi)$.

The material index is read from the object image, and the material cross section σ and density ρ are read from the material tables (see section 5.2.1.4) depending on the particle energy. The attenuation coefficient is given by $\mu = \rho\sigma$. In case a detector is found, the particle position is first translated into the detector image, as explained in section 5.2.1.1. For objects defined with Hounsfield Units (HU), the attenuation coefficient is linearly scaled between air and water or water and bone, using the bilinear conversion [Burger et al., 2002, Schneider et al., 2000]:

$$\begin{cases} \mu(HU) = \mu_{water} & HU < -1000 \\ \mu(HU) = \frac{HU+1000}{1000} \cdot \mu_{water} & -1000 \leq HU < 0 \\ \mu(HU) = \mu_{water} + \frac{HU}{3071} \cdot (\mu_{bone} - \mu_{water}) & 0 \leq HU < 3071 \\ \mu(HU) = \mu_{bone} & HU \geq 3071 \end{cases} . \tag{5.7}$$

5.2.2.5 Singles Generator

In every thread, the hit events list is read for each pair of photons in its batch of decays, analyzing the energy deposited in the detectors and in the object to generate a list of singles that are passed to the Coincidence Generator (next section, 5.2.2.6). The maximum number of singles with the same limit used for the hits events is set to guarantee that any single event is not missed. At the end of this routine, the crystal identifier of each single, its energy, its TOF, and whether the event was a scatter or a true event are saved in the global memory. We also record other output variables, such as the number of singles in each detector block, the energy histograms, or statistics about the number of hits per single event.

In figure 5.7 we show a scheme of the thread execution. At the hit events level, we analyze independently the singles generated in each photon. In case we find a hit with the object, we mark the photon as scatter event, and we skip it until we find a hit in a detector to start processing the detector signal. In the detector hits, the energy deposited in the hit and the hit TOF are blurred using the input resolution for the crystal and a Gaussian distribution. The energy resolution is scaled with the square root of the deposited energy. To mimic the Anger logic, we accumulate the total energy signal and the Anger signal given the centroid (x_i, y_i) of the crystal hit (see section 5.2.1.1, we do not add positional uncertainty in the light deposition of the scintillation crystal). The crystal position will be given by the Anger logic,

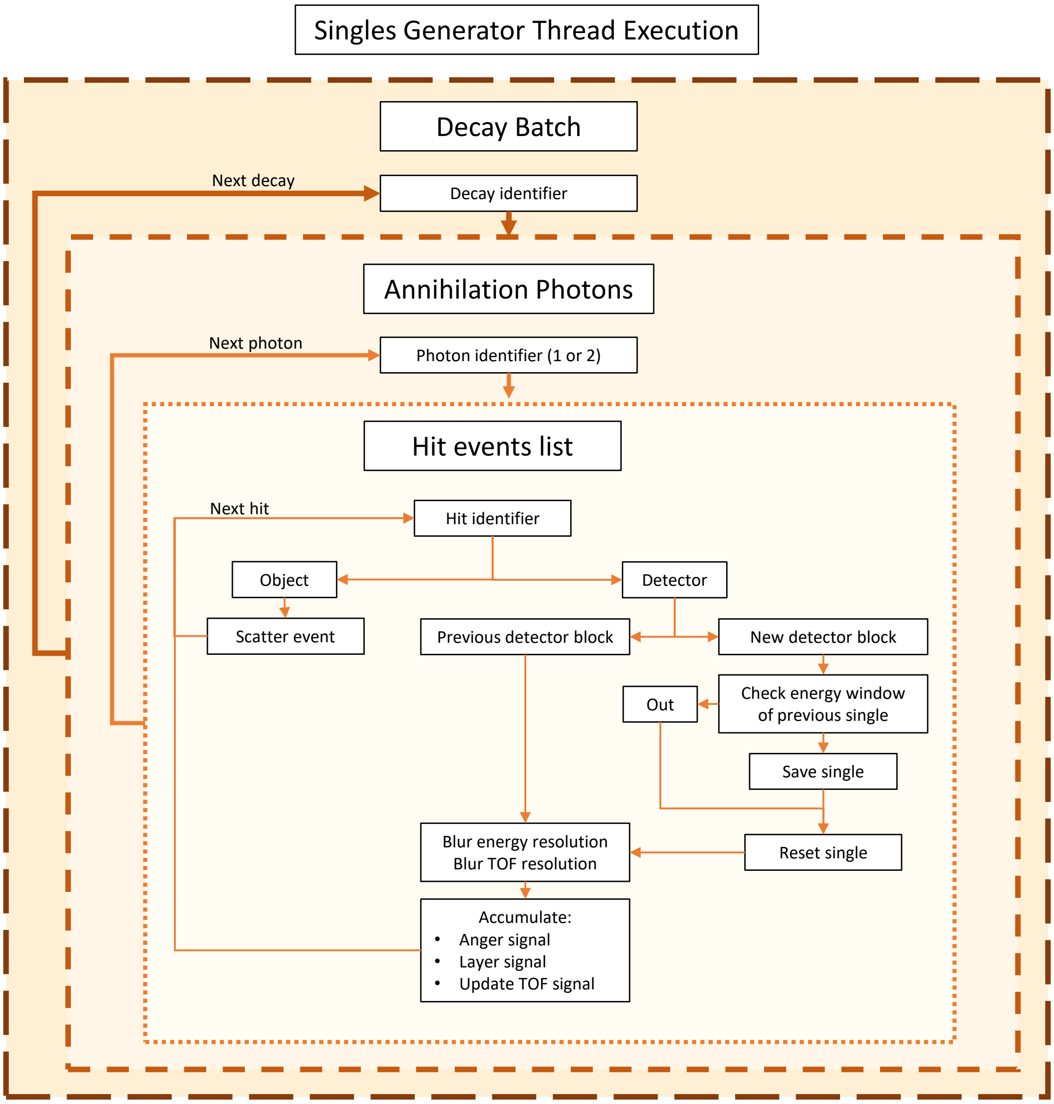


Figure 5.7: Scheme of the thread performance in the Singles Generator routine.

$$
\begin{aligned}
E &= \sum_{n=1}^{N} E_i \\
x &= \sum_{n=1}^{N} E_i \cdot x_i / E \\
y &= \sum_{n=1}^{N} E_i \cdot y_i / E.
\end{aligned}
\tag{5.8}
$$

The values (x, y) are translated to an individual crystal in the block using the look-up tables defined in the scanner inputs (5.2.1.1). The single TOF will be given by the smallest TOF signal after applying the TOF resolution. If multi-layer detectors are defined, we use a similar energy weighted method to determine in which layer the single is positioned.

The singles energy and crystal identifier are saved if the event is inside the energy window, when the hits list is finished or another detector is reached. In the latest case, we reset all the single parameters (energy, Anger signal, and timing), and the whole process is repeated. Each thread will process the hit events of at least the two photons of one decay, repeated for the whole decay batch.

5.2.2.6 Coincidence Generator

The Coincidence Generator checks the detectors involved in the singles list for every decay in its batch to determine the coincidence LOR. If the detector couple are not defined in the coincidence matrix (see section 5.2.1.1), or more than 2 single events occurred (i.e. a multiple coincidence), the event is discarded. If we accept the coincidence, the LOR histogram of true or scatter events accumulates a coincidence in the LOR index given by the detector crystals. A list of coincidences with the TOF signal (time difference between the two singles in coincidence), the singles energies, and crystal indices can be configured. This routine also generates an energy histogram for the singles that generated a coincidence event, a TOF histogram, and a detection image from the decay positions in the initialization routine (5.2.2.3) that have been detected.

5.2.3 Output files

The UMC-PET code was designed to satisfy its developers' needs in image reconstruction and scanner performance simulation, and thus the main outputs are not standardized. We must remark as well that it does not produce a closed output format. In this section we summarize a few of the usual outputs generated, such as the LOR histogram of coincidences, different list files, and other files related to general knowledge about the simulation.

- LOR histograms

  The LOR histograms match the data format used in the SuperArgus scanner [Udias et al., 2018]. In the histograms we accumulate all the coincidences at the end of every simulation step for each couple of crystal pixels in coincidence, thus avoiding any data compression. The UMC-PET separates the events marked with the scatter flag from the prompt coincidences, and it generates separate histograms for each one.

- List files

  To keep most of the information about the simulation, the code may write list files with all the details needed. The list of events matrices are loaded from the device memory to the host memory after each call to any of the main routines (see fig. 5.1), and the host CPU processes the matrices sequentially, discarding the empty spaces while saving the events in a bigger buffer (notice we work with batches of particles with an independent mini-buffer of hits, singles, or coincidences for each particle). Every time the buffer in the host is full, the code writes the list of events and resets the buffer, thus reducing the disk writing operations. The list format is tailored manually in the code depending on the experiment requirements, and it may include information about the time-of-flight, energy, scatter information, crystal index, emission voxel, or any other parameter of interest.

- Emission, detection, energy, and hits images

  To assess the UMC-PET performance at a glance, a few images with cumulative information of the simulated events are also generated. The emission image and the detection image match the source image size, and they both follow the distribution of the blurring image from the PR kernel. In the case of the detection image, only the detected emission events are saved, thus its distribution is multiplied by the sensitivity map. The hits and energy images match the scanner image size, and they respectively keep all the interactions and energy released in either the scanner, patient body, or any other object.

- Other files

  Although the output can be adapted, the default options offer information related to computation time of each section of the code, energy histograms, sensitivity, scatter fraction, and statistics about hits, singles, and coincidences.

5.3 Validation

In this section we compare the results from the UMC-PET with PeneloPET performance, using simulations of the 6R-SuperArgus scanner (see section 2.2.1). In all the PeneloPET simulations presented below, the time evolution has been idealized to avoid discrepancies between PeneloPET and UMC-PET due to the random events, pile-up events, triple coincidences, etc. We set a symbolic low activity and long acquisition time thus we can neglect the effects in the simulation due to timing effects.

5.3.1 NEMA Scatter Fraction phantoms in the 6R-SuperArgus scanner and general parameters of the simulation

For the 6R-SuperArgus scanner, we simulated the rat-like and mouse-like phantoms described in the NEMA NU 4-2008 protocol for scatter fraction (SF) assessment [National Electrical Manufacturers Association, 2008]. These phantoms are composed of a polyethilene cylinder with a line source filled with a solution of either ^{18}F or ^{11}C placed a few centimeters off-axis, parallel to the cylinder central axis. In the simulations, we are going to use water instead of polyethilene.In figure 5.8 we show a scheme of the simulation phantoms inside the scanner. The mouse-like phantom has 25 mm diameter and 70 mm length, and the source is 10 mm off-axis and 60 mm long, whereas the rat-like phantom has 50 mm diameter and 150 mm length, and the source is 17.5 mm

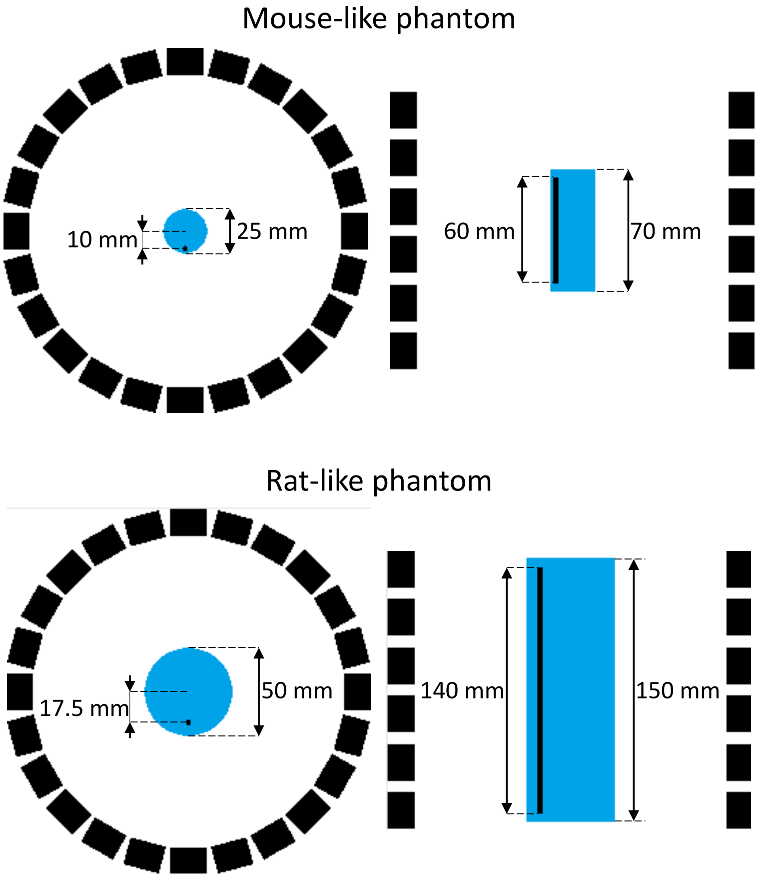


Figure 5.8: Scheme of the simulated NEMA NU 4-2008 mouse-like (up) and rat-like (down) phantoms for scatter fraction assessment. For both phantoms we show a central slice of the transverse view (left) and sagittal view (right).

Image	**#Voxels**	**Voxel size** (mm)	**Volume size** (mm)	**Data format**	**Memory size** (MB)
Scanner & object	269× 269× 208	0.775× 0.775× 0.775	208.5× 208.5× 161.1	Signed Short Integer (2B)	28.7
Detector	13× 13× 15	1.55× 1.55× 1	20.15× 20.15× 15	Signed Short Integer (2B)	4.95 (kB)
Source (Mouse)	60× 60× 100	0.1× 0.1× 0.12	6× 6× 12	Float (4B)	1.37
Source (Rat)	60× 60× 100	0.1× 0.1× 0.28	6× 6× 28	Float (4B)	1.37

Table 5.2: Details of the voxelized volumes used in the UMC-PET simulator. Notice that scanner and object were the same for both phantoms, but they had different voxel size for the source image.

off-axis and 140 mm long. In both cases the line source has a diameter of 3 mm, and ^{18}F was selected for the PR.

The scanner geometry, dimensions, and other characteristics such as energy and time resolution were taken from the published information, explained in section 2.2.1. In the next section 5.3.3, we compare the results for different energy windows. The reflectors between the detector pixels (0.05 mm in the official dimensions) were removed since the UMC-PET would require a pixel size smaller or equal to the reflector size in the detector image. This is not a limitation for the code, but we would rather use smaller images to improve the performance. It is important to remark that the reflectors major impact is in the scanner sensitivity, and this can be easily accounted for. In table 5.2 we list the image details declared for the simulations in the UMC-PET.

In all the simulations of the UMC-PET a total number of decays of $6.228{\cdot}10^7$ were simulated, equivalent to 18 minutes of acquisition with 100 µCi total activity of ^{18}F, that are standard activity and acquisition time values in preclinical PET studies. In the PeneloPET simulations, we set the activity to 10^4 Bq (0.27 µCi) and the acquisition time to 10^4 seconds (167 minutes).

To estimate the SF, the NEMA NU 4-2008 protocol [National Electrical Manufacturers Association, 2008] indicates that the acquisition data must be sinogrammed using the single slice rebinning (SSRB, see section 1.3.2.4), discarding all the LORs further

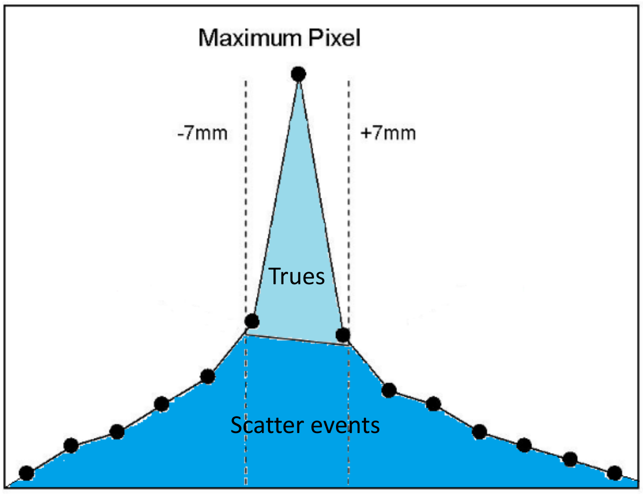


Figure 5.9: Scheme of the profile in the NEMA Scatter Fraction phantoms sinogram with SSRB, maximum pixel aligned in the center of each projection, and angular-axial projection collapsed (image modified from [National Electrical Manufacturers Association, 2008]). The dark-blue region defines the scatter area, whereas the light-blue region is the true events area.

8 mm from the edge of the phantom. Every projection angle is displaced to align the pixel of maximum intensity in the center of the projection, and the resulting sinogram is collapsed angularly and axially to obtain a summed profile (see figure 5.9). The NEMA protocol considers that all the events farther than a central strip of 14 mm are scattered and random events, besides the region below the line connecting the profile intersection with the 14 mm strip. The intersection values with the strip are calculated using linear interpolations, averaged, and multiplied by the number of points inside the strip (including fractional values) to estimate the background contribution below the trues region. The total scatter events are the sum of the events outside the strip and the background inside the strip. The SF is given by the ratio between the total events and the scatter events.

5.3.2 CPU and GPU devices

In the PeneloPET simulation, we ran all the codes in a single core using a Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz. For the simulation of UMC-PET, we implemented the code in a Intel(R) Xeon(R) W-2155 CPU @ 3.30GHz for the host computation, and

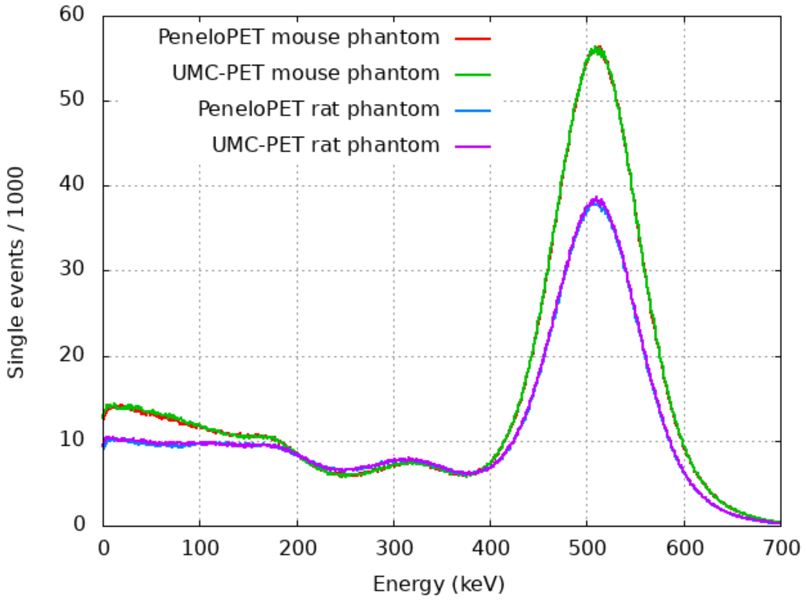


Figure 5.10: Energy histograms generated by the UMC-PET and PeneloPET for the mouse-like and rat-like phantoms without energy window.

a 11 GB GeForce RTX 2080 Ti GPU, with 4352 cores, for the parallel computation in the device.

5.3.3 Results

Figure 5.10 shows the energy histograms generated by UMC-PET and PeneloPET for both phantoms in figure 5.8. The energy histograms have been generated for the coincidence events of acquisitions without energy window to ensure the code accuracy for any energy left in the detector.

We show the collapsed projection profiles calculated as aforementioned in 5.3.1 and figure 5.9 for both simulators for the mouse and rat phantoms in figure 5.11. In table 5.3, we show the SF calculated for all the profiles using the NEMA protocol, and the SF and sensitivity given by the simulation output information. The NEMA SF shows great agreement among the different tests: 1.22%, 0.834% for the mouse phantom, 100-700 keV and 425-600 keV energy windows respectively, and 0.992% and 2.79% for the rat phantom, 100-700 keV and 425-600 keV energy windows respectively. The SF given by the simulators shows higher discrepancies: 6.94%, 5.55% for the mouse phantom, 100-700 keV and 425-600 keV energy windows respectively, and 6.61% and 5.64% for the rat

Mouse-like phantom						
	NEMA SF (%)		**Simulated SF** (%)		**Sensitivity** (%)	
keV	100-700	425-600	100-700	425-600	100-700	425-600
UMC-PET	7.39	4.47	14.72	5.64	6.45	3.03
PeneloPET	7.48	4.51	13.70	5.33	6.32	2.96

Rat-like phantom						
	NEMA SF (%)		**Simulated SF** (%)		**Sensitivity** (%)	
keV	100-700	425-600	100-700	425-600	100-700	425-600
UMC-PET	24.69	11.21	33.49	11.93	5.03	1.93
PeneloPET	24.93	11.52	31.28	11.26	4.93	1.89

Table 5.3: Estimated SF using the NEMA protocol, SF from the simulations, and sensitivities for the mouse (up) and rat-like (down) phantoms using UMC-PET and PeneloPET.

phantom, 100-700 keV and 425-600 keV energy windows respectively. It is important to highlight that the errors on the simulated SF must be caused by different definitions in the code, since the measured differences in the NEMA SF are much lower. Finally, the sensitivity shows acceptable discrepancies: 2.02%, 2.35% for the mouse phantom, 100-700 keV and 425-600 keV energy windows respectively, and 2.03% and 1.97% for the rat phantom, 100-700 keV and 425-600 keV energy windows respectively.

In table 5.4, we show the computation time of each simulation, comparing the decays and coincidences generated per second. We observe an acceleration factor of ∼2000 in all the simulations when the UMC-PET simulator is used. The simulations performed with the UMC-PET required and additional time of ∼3 seconds to load all the input files into GPU memory, but we did not account for it to estimate the decay and coincidence rate since this is a constant time, independent of the number of simulated decays. Compared to the simulation time, we have neglected the loading time of the input files in PeneloPET. The time required for each main routines in the UMC-PET simulator is presented in table 5.5, showing that the simulator spends most of its time in the photon tracker. The remaining time is devoted to CPU general operations (<1%). PeneloPET output did not provide the time allocation of each part of the code, thus no comparison against individual routines is provided.

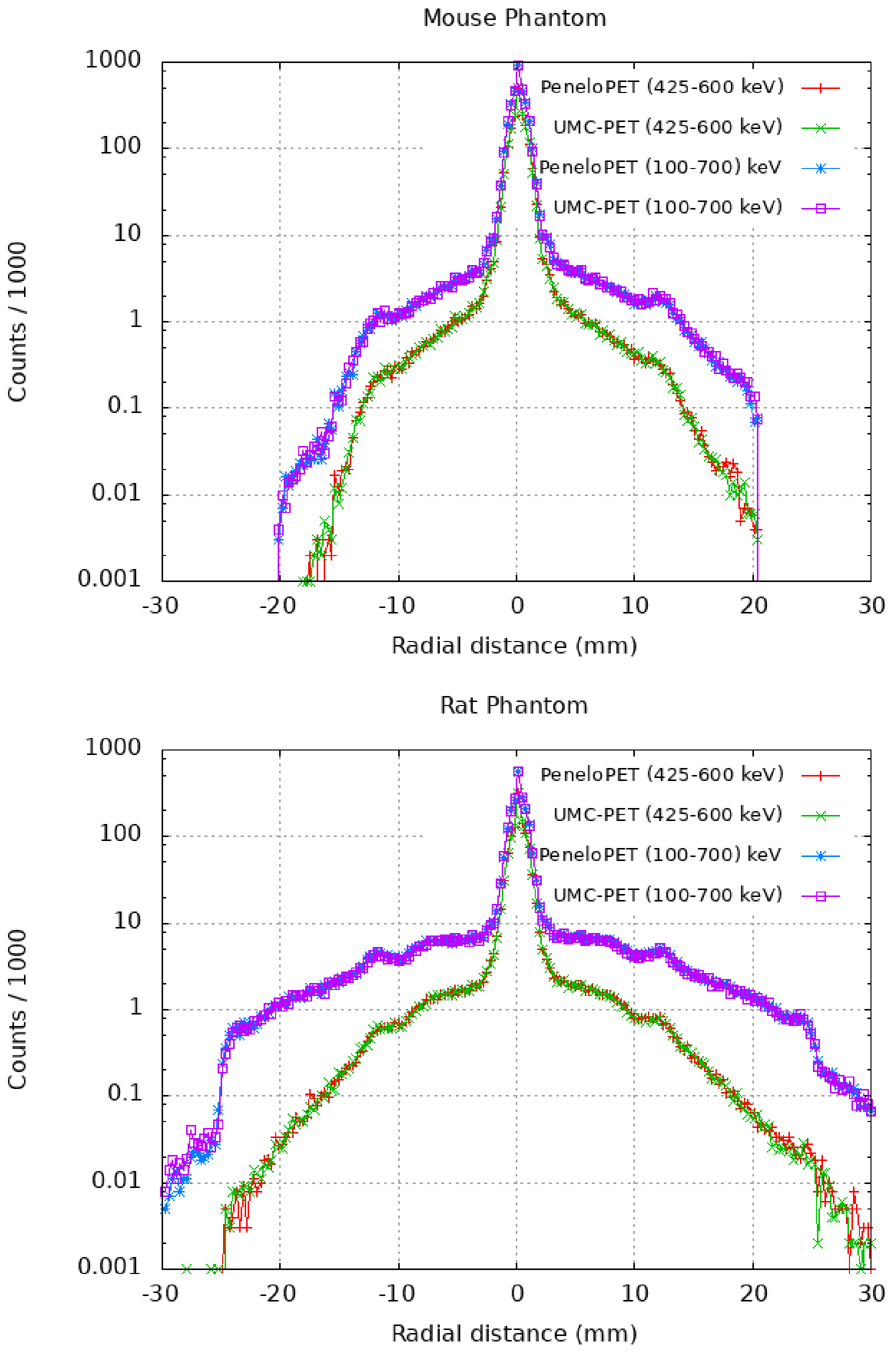


Figure 5.11: Angular-axially collapsed and maximum-centred profiles of the projections simulated with the mouse (up) and rat-like (down) phantoms for two different energy windows (425-600 keV and 100-700 keV) for PeneloPET and the UMC-PET simulator.

Mouse-like phantom						
	Simulation time (s)		**Decays/s**		**Coincidences/s**	
keV	100-700	425-600	100-700	425-600	100-700	425-600
UMC-PET	1.33	1.31	$4.68{\cdot}10^{7}$	$4.76{\cdot}10^{7}$	$3.00{\cdot}10^{6}$	$1.43{\cdot}10^{6}$
PeneloPET	3035	2845	$2.05{\cdot}10^{4}$	$2.19{\cdot}10^{4}$	$1.30{\cdot}10^{3}$	$6.47{\cdot}10^{2}$

Rat-like phantom						
	Simulation time (s)		**Decays/s**		**Coincidences/s**	
keV	100-700	425-600	100-700	425-600	100-700	425-600
UMC-PET	1.91	1.90	$3.25{\cdot}10^{7}$	$3.29{\cdot}10^{7}$	$1.63{\cdot}10^{6}$	$6.29{\cdot}10^{5}$
PeneloPET	3664	3817	$1.70{\cdot}10^{4}$	$1.63{\cdot}10^{4}$	$8.39{\cdot}10^{2}$	$3.08{\cdot}10^{2}$

Table 5.4: Computation time (simulation time, number of simulated decays per second, and coincidences generated per second) for the mouse (up) and rat-like (down) phantoms using UMC-PET and PeneloPET. The values presented in this table for the UMC-PET did not account for the time needed for input reading and data loading in the GPU (∼3 seconds for each simulation). The details on the platforms for each simulator are specified in section 5.3.2.

Computational time in UMC-PET									
Phantom	keV	**Decay init.**		**Photon tracker**		**Singles Generator**		**Coinc. Generator**	
		(s)	(%)	(s)	(%)	(s)	(%)	(s)	(%)
Mouse	100-700	$2.33{\cdot}10^{-2}$	1.77	1.12	85.1	0.15	11.3	$2.00{\cdot}10^{-2}$	1.5
	425-600	$2.34{\cdot}10^{-2}$	1.81	1.12	86.7	0.13	10.2	$1.39{\cdot}10^{-2}$	1.1
Rat	100-700	$2.34{\cdot}10^{-2}$	1.21	1.73	89.7	0.15	7.9	$1.95{\cdot}10^{-2}$	1.0
	425-600	$2.41{\cdot}10^{-2}$	1.27	1.73	90.8	0.13	7.1	$1.31{\cdot}10^{-2}$	0.7

Table 5.5: Computational time of the main routines (photon tracker, singles generator, and coincidences generator) in the UMC-PET simulator. The total time is given in table 5.4.

5.4 Scatter correction

Scatter correction in PET is a well-known need since up to a 50% of the detected prompt events may have undergone scatter in a clinical acquisition [Adam et al., 1999, Ye et al., 2011] (for small animal PET scanners, the impact is noticeably reduced), and non-corrected images suffer from biased uptake and artifacts that reduce the quality of quantitative assessment [Afshar-Oromieh et al., 2017, Heußer et al., 2017, Lindemann et al., 2019].

Traditionally, the scatter contribution to the data is either subtracted or added to the projection model. The scatter distribution depends on the activity distribution and the anatomical structures of the patient, thus requiring input images of both the emission map and the µ-map to estimate it. The most commonly used method in PET for scatter estimation is the single scatter simulation (SSS) [Ollinger, 1996, Watson, 2000, Watson et al., 1996], a fast and analytical algorithm that is based on the probability of having a coincidence event that suffered a single Compton scatter in either one of its photons. However, the SSS has intrinsic drawbacks. The method does not estimate multi-scatter events, that may account from 15 to 40% of total scattered events [Adam et al., 1999, Ye et al., 2011] and can result in wrong contrast values [Polycarpou et al., 2011], it does not model the detectors (in [Ma et al., 2020], the authors stressed the relevance of detector modeling in scatter correction), and it requires for a global scaling factor for the SSS distribution. This factor is usually obtained using tail fitting strategies in the scatter-only regions of the sinograms [Thielemans et al., 2007, Werling et al., 2002] that may result in noticeable halo artifacts around high organ-to-background ratio organs, like the bladder or the kidneys [Heußer et al., 2017, Lindemann et al., 2019].

Monte Carlo simulations may include all the missing physics in the SSS algorithm, besides intrinsically scale the scatter estimation with great accuracy [Castiglioni et al., 1999, Holdsworth et al., 2001, Kyung Sang Kim et al., 2014, Ma et al., 2020]. Some authors combined MC simulations and the SSS algorithm to improve the algorithm robustness with a global scaling factor obtained from a short simulation of the MC model [Pönisch et al., 2003, Ye et al., 2014]. MC methods were combined with variance reduction techniques to increase the computational performance [Castiglioni et al., 1999, Holdsworth et al., 2001], but it was not until GPU parallelization that they were considered for practical use [Kyung Sang Kim et al., 2014, Ma et al., 2020]. To the best of our knowledge, only GGEMS [Ma et al., 2020] has been applied so far for pure MC scatter estimation accounting for the detector modeling as well. Other authors already tried to improve the MC performance with DL in either the projection space [Berker et al., 2018, Laurrent et al., 2020, Qian et al., 2017, Xiang et al., 2020], or directly in the image domain [Arabi et al., 2020, Shiri et al., 2020, Yang et al., 2019, 2021].

In this section we show a proof of application of the UMC-PET simulator for scatter correction.

5.4.1 Methods

5.4.1.1 Human study in the Biograph mMR scanner

We applied the UMC-PET scatter estimation to a human study of a female of 57 years old and 75 kg, injected with ^{18}F-FDG (fluorodeoxyglucose) acquired in the Biograph mMR scanner (see section 2.2.2) with two bed positions centered in the pelvis (first) and thorax (second). The subject gave written informed consent, and the local Institutional Review Board approved the study. The patient was injected 389 MBq (10.5 mCi) 1 hour and 23 minutes before the first bed acquisition. Each acquisition took 14 minutes, having $2.87 \cdot 10^8$ and $2.93 \cdot 10^8$ total coincidences respectively. The data provided by the scanner use the Siemens sinogram format, with axial compression of span 11 and maximum ring difference of 60, resulting in 11 segments (1 direct, 5 of positive inclination, and 5 of negative inclination) with a total of 837 sinogram slices. The scanner provided the prompts sinogram, the randoms sinogram, the normalization sinogram, the SSS sinogram, and the µ-map in Hounsfield units. The µ-map does not include the arms of the patient, thus we expect misscorrection in the regions where the anatomical information is missing [Afshar-Oromieh et al., 2017].

For the scatter correction, we compared the SSS estimation provided by the scanner combined with the scatter fraction computed by the UMC-PET (a similar approach was used in [Ye et al., 2014]), and the pure UMC-PET estimation. In both cases, we ran a reconstruction without scatter correction (including normalization, attenuation, and random corrections) that was used as input in the UMC-PET simulator, together with the µ-map. To account for the scatter events outside the FOV, we used the whole image (both beds) as decay map, although the scatter was simulated separately for each bed. The voxel size used for the source was of $2 \times 2 \times 2.03$ mm^3 (the same size used in the image reconstruction), and the sizes in the µ-map were interpolated from $2.6 \times 2.6 \times 3.12$ mm^3 to $2 \times 2 \times 4$ mm^3 for the simulation. The UMC-PET simulation used to obtain the scale factor for the SSS took less than 30 seconds since we only needed the ratio between prompts and scatter events. For the pure UMC-PET estimation, we used a longer simulation of 2500 seconds (same system used in section 5.3.2). Since the scatter distribution is smooth, we averaged the events with closest neighbouring LORs to reduce the variability of the UMC-PET scatter estimation.

We used the GPU-based reconstruction software presented in chapter 3 with the normalization and randoms corrections provided by the scanner, and the attenuation corrections obtained after the µ-map projection. The sinograms were adapted to match the LOR histogram format. The images were reconstructed with 4 iterations of 5 subsets of the OSEM algorithm. The final images of each bed were combined with an overlapping of 30 slices (6.09 cm).

The reconstructed image for this patient was also provided by the scanner. In this case, 3 iterations and 21 subsets were used. To assess similar image quality in both methods, we measured the noise inside the liver and stoped our reconstructions at a similar noise level.

5.4.2 Results

In figure 5.12 we show a radial-angular slice of the scatter sinograms with SSS and UMC-PET scale, and with pure UMC-PET prediction. A radial profile is shown in the botton of the image. In figure 5.13, a similar example is shown with an axial-radial plane of the sinograms data set. The line profile compares the differences in the axial direction. In both figures we can notice the difference in the scatter distribution between the SSS and the UMC-PET. We do not know the actual scale of the SSS used for the scanner images, thus we can only compare with the UMC-PET scaling. Higher variance due to the MC noise can be observed in the profiles of the UMC-PET estimation.

Figure 5.14 shows the reconstructed image provided by the scanner, the reconstructed image using the scatter from the SSS scaled with the UMC-PET, and the reconstructed image with the pure UMC-PET scatter correction. An elliptical region inside the liver has been drawn to find noise equivalent images. The measured noise (defined as $100 \cdot std.dev./mean$) was 14.2% in the Biograph mMR image, 12.8% in the SSS scaled with the UMC-PET, and 12.2% with the pure UMC-PET scatter correction.

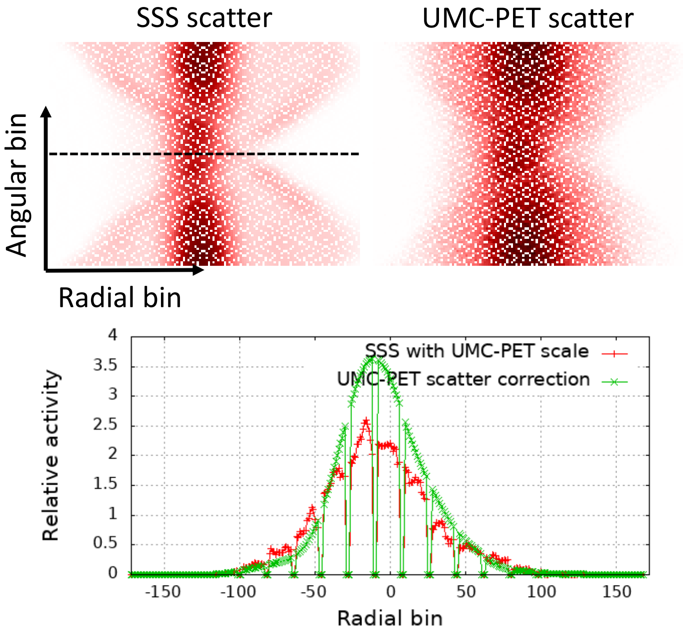


Figure 5.12: Sinogram slice of the SSS with UMC-PET scale, and the UMC-PET scatter sinograms in the pelvis-centered bed. The line profile shows the differences between both estimations at different radial bins. The sinogram slice corresponds to a direct sinogram above the bladder position.

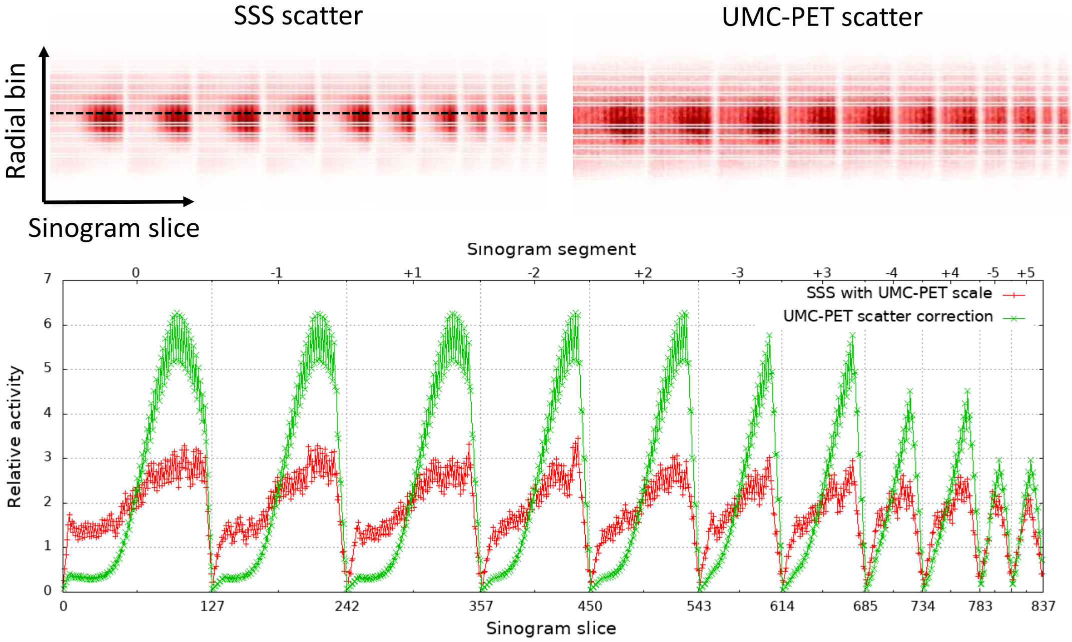


Figure 5.13: Axial-radial sinogram view of the estimated scatter obtained with SSS and with UMC-PET scale, and the UMC-PET scatter estimation. The sinograms in the pelvis-centered bed are shown. The line profile shows the differences between both estimations at different sinogram slices. The segments can be visually identified in the images, and they have been depicted in the upper axis of the profile with positive and negative indexes defining the sinograms inclinations. The profile corresponds to a section through the bladder.

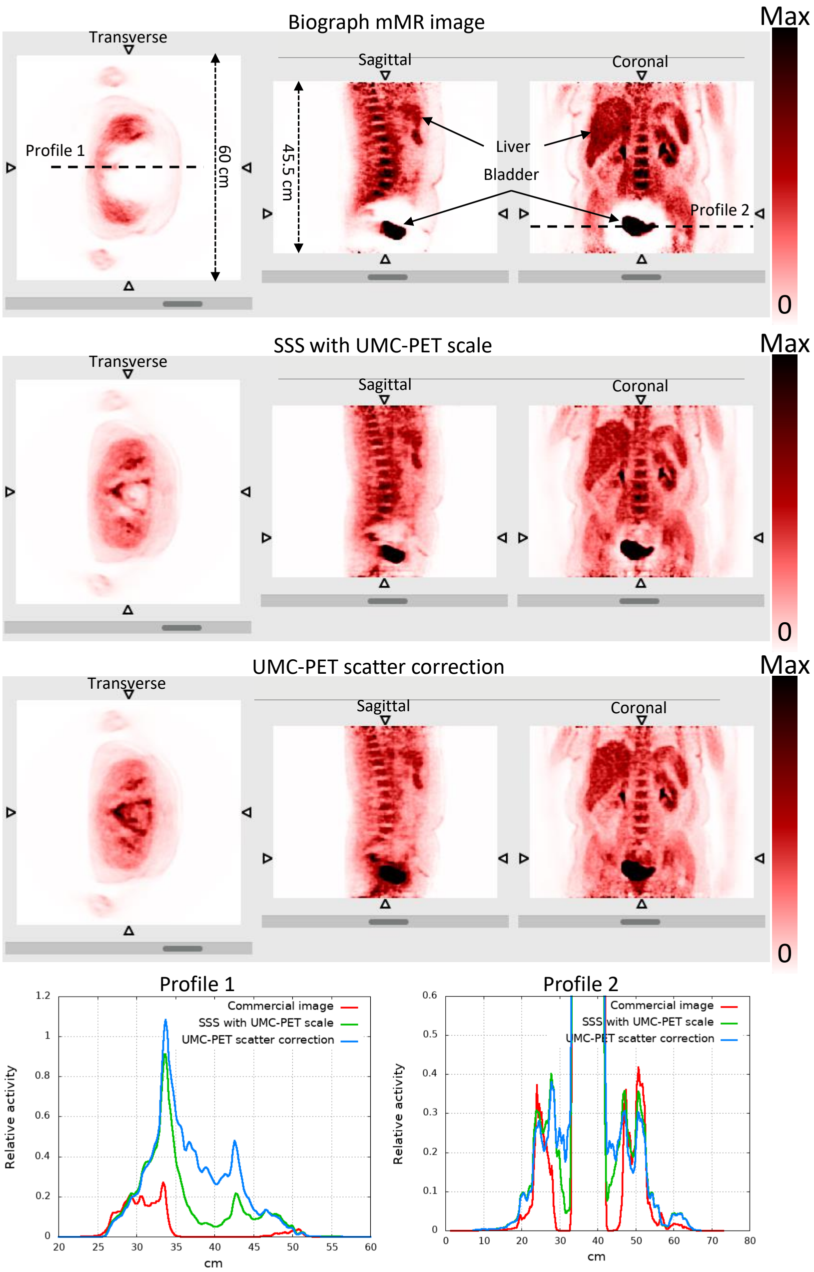


Figure 5.14: Reconstructed PET image of a human patient acquired in the Biograph mMR scanner (up), reconstruction with the scanner SSS scaled using the UMC-PET simulator (middle), and reconstruction with the UMC-PET scatter estimation. In the bottom of the image two line profiles are shown. The image scale has been scaled up to twice the activity in an ROI inside the liver.

To obtain equivalent color scales in all the images, we also used the region inside the liver to define the color range from zero to twice the mean activity inside the liver. The regions where the uptake is higher than this value are shown in black (e.g. the bladder). We can notice that the halo artifact around the bladder is diminished when we introduce the UMC-PET scale in the SSS, and it is completely absent with the UMC-PET scatter correction. Below the shoulders, the attenuation correction of the arms cannot be estimated since the anatomical information is missing in the µ-map. For this reason, the arms show less uptake than expected.

5.5 Discussion

In this chapter we have presented the Ultra-fast Monte Carlo PET simulator (UMC-PET). Among the wide range of PET dedicated and general MC codes we mentioned in the introduction, the Ultra-fast Monte Carlo PET simulator finds is place thanks to the GPU acceleration and a modular design. Compared to other GPU PET dedicated codes [Lai et al., 2019, Ma et al., 2020], our implementation offers a different approach with both advantages and (minor) drawbacks. We tried to develop the simplest code that could attain image reconstruction assistance and scanner design, while optimizing three main features: speed, flexibility, and accuracy.

The UMC-PET speed was achieved through GPU execution. The most computationally demanding routine was the Woodcock routine for photon tracking (see table 5.5). For the sake of simplicity, we reduced the parallelization problem of particle tracking to a one-story-per-thread approach. Although some authors discussed about the need of optimization schemes to avoid thread divergence in the GPU [Hissoiny et al., 2011, Jia et al., 2011], this conflict was not a problem for simulating gamma particles alone. The thread divergence is caused by multiple threads performing different paths in a warp (e.g. because of a non-deterministic process, such as those given in MC simulations), where all the threads are grouped by the GPU multiprocessor to perform instructions in a warp lockstep. This architecture is called single-instruction multiple thread (SIMT) [Fatica and Ruetsch, 2014]. In terms of execution for our software, this means that threads in a warp performing too different paths will be serialized. However, this would be an issue to simulate different particles per thread (i.e. very different paths performing at different threads), or for different threads with very different lifetimes (in terms of particle steps) [Hissoiny et al., 2011]. In our case we are considering one particle type (gamma photons), but we could find divergence caused by different kind of interactions or stories with different lifetimes, although this fact was not reported so far in the literature. In the future we will study approaches to reduce thread divergence.

The one-story-per-thread implementation in the GPU also simplified the code, allowing the factorization of the simulation in the different main routines of section 5.2.2. Since the primary purpose of UMC-PET was to assess image reconstruction, we removed the time evolution from the main routines, avoiding sequential calculations (such as coincidence sorting) that are impractical for optimal GPU parallelization. Time of flight (TOF), however, is fully incorporated at each history, hit, or single. In GGEMS [Ma et al., 2020], the time evolution is implemented after each batch of decays, adding a global time stamp to each decay and the TOF for each gamma photon, and gPET [Lai et al., 2019] parallelized the time evolution using a sophisticated adding scheme for pulses generated from photons that are tracked in different threads, with time sorting. In UMC-PET we can generate a list of hit events or single events that is post-processed adding a global time stamp in each decay and any effect that involves the timing evolution. This strategy permits using the same list of events for different activity rates, gaining flexibility to estimate single rates, coincidence rates, random coincidences, pile-up events, or other time-dependent events.

In section 5.2.2.2, we detailed the approximated PR blurring kernel implemented to describe the annihilation distribution based on the decay distribution and the density map. The PR kernel is optionally applied prior to the simulation, thus eliminating the positron simulation to reduce the computational burden and simplify the code. Since the parameters fitted in the table 5.1 to the analytical formula defined in equation 5.3 were obtained for the distribution of the annihilation events in water, the PR was approximated for other materials using the electronic density. Cal-Gonzalez [2014], Cal-González et al. [2015b] benchmarked this method against simulations, showing accurate results in homogeneous media. However, the error introduced by the PR kernel will increase at the interfaces of different materials. Other authors have increased the accuracy of PR kernels using deep learning techniques and convolutional neural networks trained with simulated images, like Herraiz et al. [2020] already did to transform ^{68}Ga images in equivalent ^{18}F images. It is important to remark that there are a lot of uncertainties among the simulation of the positron range of different authors [Cal-González et al., 2013], thus the approximation considered in this work satisfied the purposes of the UMC-PET simulator.

Another speed-up improvement is achieved with the random selection of the scattering angles and energies from a precomputed sample (see section 5.2.1.4). In general, the samples from different materials at different energies will be mixed, preventing poor statistics while reducing the need for calculations on-the-fly. The results in 5.3 showed good enough accuracy.

The voxelized scheme presented in section 5.2.1.1 is easily configurable to any scanner geometry or crystal pixelization. So far, the 6R-SuperArgus has been presented in section 5.3, but non-cylindrical scanner geometries can also be implemented. Spherical geometries are gaining relevance among brain dedicated PET scanners [Catana, 2019b, Chaudhari and Badawi, 2021, Tao et al., 2018, Yoshida et al., 2020] and small animal scanners [Perez-Benito et al., 2018] due to the enhanced sensitivity. In 2020, a simulation of a prototype of the HSTR-BrainPET with spherical symmetry was presented in [Galve et al., 2020b]. These results might be found in chapter 3. Other non-standard approaches that could be implemented with the UMC-PET include PET inserts or non-traditional shapes to gain both resolution and sensitivity [Akamatsu et al., 2021, Grkovski et al., 2015, Qi et al., 2011, Samanta et al., 2021], complex multi-layered detectors [Mohammadi et al., 2017, Wang et al., 2006], or non-prism shaped pixelated crystals [Perez-Benito et al., 2018].

On the other hand, we separated the voxelization of the scanner image and the block image (see section 5.2.1.1 and image 5.3) to reduce the memory needed. Although current GPUs are capable to fit large arrays in the global memory (up to several gigabytes), the voxelized approach may seem impractical from a memory point of view since large images may be required to define the whole scanner in detail, which will limit the voxel size, besides large empty spaces inside the scanner FOV have to be kept in memory. Parametrized surfaces and volumes would reduce considerably the memory requirements, but it would increase the amount of calculations on-the-fly. So far, the two-level factorization of the scanner shows a good trade-off between computation performance and memory usage under the GPU limits.

The accuracy of UMC-PET was validated against PeneloPET in section 5.3. The results shown in figure 5.10, figure 5.11 and table 5.3 show good agreement between both simulators. The largest error is obtained for the simulated SF, but this discrepancy might be caused by differences on determining the scatter events by each simulator. The NEMA SF, which evaluates the distribution of the coincidences inside the scanner shows the smallest variations among the simulations. We assume that the discrepancies in the simulators performance is caused by differences in the code scheme, like the voxelized representation of the volumes in UMC-PET against the parametrized volumes of PeneloPET, or different implementations of the detector response. In any case, these are minute differences. In further studies we will validate the UMC-PET simulator using a clinical scanner. We expect that the errors scale with the scanner size, without being remarkable.

UMC-PET parallel implementation in the GPU clearly outperformed PeneloPET (see table 5.4), obtaining a speed factor of ∼2000×. It is important to mention that PeneloPET (and any CPU-based simulator in general) can be easily paralyzed in multiple CPU cores, achieving an acceleration equivalent to the number of cores (common computers use 4 or 8 cores, thus the acceleration gain thanks to the GPU usage is still very remarkable). GGEMS [Ma et al., 2020], gPET [Lai et al., 2019] and MGGPU-PET [Lopez-Montes, 2021] already demonstrated speed gains when the PET simulation was taken to the GPU with different schemes, but there is not reliable comparison among these GPU simulators in similar platforms.

In both PeneloPET and UMC-PET, the simulations are faster for smaller phantoms (see the time comparison of table 5.4). From the results in table 5.5 we can infer that this is related with the time spent in the photon tracker. The photons suffer a larger number of interactions in larger phantoms, increasing the number of steps of the particle and the computational cost of this routine. Further, besides phantom size, we have observed that simulations of larger physical volumes require a higher computational time (for example, simulations of clinical scanners) because the Woodcock algorithm (section 5.2.2.4) is going to take a higher average number of steps until a photon reaches a detector or the end of the simulating volume, increasing the computational cost of this routine. In this case, due to steps in empty areas of the scanner.

In section 5.4, we have implemented the UMC-PET simulator for scatter correction in the reconstruction process. We found noticeable differences between the Biograph mMR SSS and the UMC-PET scatter distributions in the projection space (see figures 5.12 and 5.13). Scatter estimated from the scanner software shipped with the mMR resulted in a halo around high contrast regions, like the bladder. The halo effect might be mitigated computing the appropriate scatter scale with MC methods, as it has been shown in figure 5.14 (this technique was already proved by Ye et al. [2014]). The structures hidden by the halo are recovered when the UMC-PET scatter is used. In the upper half of the sagittal and coronal planes of figure 5.14 (second bed of the study, centered in the thorax), where the uptake contrast among the organs is smaller, all the scatter corrections show clear images. Although we have observed differences in the scatter profiles of the estimated sinograms in this regions as well, the differences are not visually noticeable. The differences between the SSS and the UMC-PET scatter distribution may be due to multi-scatter events, the detector modeling, or differences in the estimation of scatter cross sections and angles. It is also important to remark that the µ-map presented in the study of section 5.4.1.1 did not include the arms of the patient, thus underestimating the attenuation and scatter corrections in all the cases, and this might be partially related to the halo artifact [Afshar-Oromieh et al., 2017].

The results shown in section 5.4.2 were obtained with a simulation that took 2500 seconds in a single RTX 2080 Ti and a mean filter applied to closest neighbours in the scatter distribution. The simulation time could be reduced with multiple GPUs or with faster models (such as the RTX 3090). The variance in the UMC-PET estimation might be reduced using more powerful filters (like a Gaussian filter), exploiting the scatter smoothness. Deep learning techniques can be implemented here to denoise scatter estimations and increase speed.

Chapter 6

Full integrated tool: UMC-PET improved reconstruction

6.1 Introduction and motivation

We have already discussed the relevance of the System Response Matrix (SRM) in the introduction section 1.4.2.1. Usually, PET modelling is decomposed into SRM, attenuation, normalization, scatter and random components [Comtat et al., 2004, Michel et al., 2000, Zaidi and Karakatsanis, 2018]. The SRM accounts for all the effects that limit the image resolution, such as scanner geometry, positron range, photon non-collinearity, or depth of interaction (DOI) in the detector [Cherry et al., 2012, Mohammadi et al., 2019]. To obtain the most accurate results, we need to tune the model to experimental measurements or MC simulations [Iriarte et al., 2016].

Experimental methods [Gong et al., 2017, Kotasidis et al., 2011, Miranda et al., 2020, Rapisarda et al., 2010, Tohme and Qi, 2009, Zhou and Qi, 2011] measure the SRM with point source acquisitions which need high position accuracy and correction for normalization, attenuation, scatter and random coincidences. Application of these corrections is complex and further access to scanner point source measurements is not always an option. Often the results of point source measurements are combined with analytical projection models to facilitate the determination of the SRM, but these may not accurately describe the physics of the model.

When MC simulations are employed to obtain the SRM, we have a better control of the scanner response and the parameters effecting it, though we depend on the accuracy of the simulations [Gillam and Rafecas, 2016]. In many cases, hybrid methodologies combining MC simulations with analytical methods have been proposed to exploit SRM sparsity [Herraiz et al., 2006a, Li et al., 2015, Meng et al., 2021, Pilleri et al., 2019, Wei and Vaska, 2020, Xu et al., 2019, Zhou and Qi, 2014]. To avoid analytical approximations, other authors explicitly computed the elements of the SRM with extensive MC simulations [Chien-Min Kao et al., 2007, Meng et al., 2019, Rafecas et al., 2004, Zhang et al., 2010]. In all of the approaches suggested so far, the attenuation map is not considered during SRM estimation, and attenuation and scatter events are estimated separately. Often, scanner symmetries and SRM sparsity are exploited to increase the statistics in the MC simulation and to reduce the storage needs coming from the large number of LORs and voxels. In [Southekal et al., 2011], the authors implemented SimSET [Baum and Helguera, 2007] to fully simulate the SRM of the RatCAP preclinical scanner including an attenuation map to predict MC scatter and attenuation events. In this case, the SRM estimation took 10 days, so they used of a standardized attenuation map for all the reconstructions. Since the RatCAP scanner is a rat brain dedicated scanner considerably smaller than usual preclinical scanners in terms of FOV (smaller than 4 cm × 2 cm transversely-axially) and number of LORs, the effect of attenuation and scatter was not relevant, but a faster approach is needed to extend this procedure to larger scanners and general acquisitions.

In this work, the UMC-PET simulator has been implemented to accelerate the SRM estimation for standard preclinical scanners in two different ways. First, we give the details of a hybrid image reconstruction method based on a fully-MC forward projection model (UMC-PET projection) and a backward projection optimized with an analytical approach. The UMC-PET projection accurately describes all the physics involved, including photon emission, attenuation, scatter, and detection in the detectors, specifically computed for each patient attenuation map. In this case, SRM sparsity cannot be exploited because the scatter components are integrated in the simulation, and explicit storage of the SRM is impracticable, thus a list of LOR-voxel events (emission voxel and the LOR where the coincidence was detected) is used to compute the forward projection. Secondly, we use a similar approach in which the scatter and attenuation have been factorized. This method exploits the scanner symmetries to increase the statistics in the MC projection (symUMC-PET). The results of this chapter were presented in [Galve et al., 2021a,b].

This chapter is organized as follows. In section 6.2 we give the detail of the reconstruction methods and PET data acquisitions used. The results are summarized in section 6.3. The chapter is closed with the discussion of this work in section 6.4.

6.2 Methods

In this section we explain all the reconstruction methods proposed for this chapter and the data utilized to assess the performance of the reconstruction. We will use the GPU-PET reconstructor from chapter 3 as reference.

6.2.1 The GPU-PET reconstructor with spatially variant PSF

The GPU-PET reconstructor from chapter 3 was modified to accept spatially-variant PSF (SV-PSF) in the image blurring kernel (section 3.2.1.1). The PSF was approximated by a 3D-gaussian with components in the radial, tangential, and axial directions, measured at different positions of the scanner from MC simulated acquisitions of point sources every 5 mm. The radial and tangential profiles were evaluated along the radial direction in the central transverse plane. The axial profile was evaluated along the axial direction in the central axis of the scanner. The Gaussian FWHM is interpolated at every position of the FOV from the measured values. The PSF blurring kernel performs separately the convolution on the transverse plane (mixing up the radial and tangential PSF) and the axial direction.

We thank Fernando Arias Valcayo for helping with these modifications on the code.

6.2.2 The UMC-PET integration with the MLEM

We have implemented the UMC-PET simulator from chapter 5 as a projector. In the MLEM equation 1.20, this means that the estimation of the projected data $\tilde{y}_i$ is based on a realistic simulation. In practice, we computed a list of detected events with the emission voxel for each event and the LOR index where it was detected. This list is used to obtain $\tilde{y}_i$. The backprojection was implemented using the same scheme implemented in the GPU-PET reconstructor, decomposed in the blurring kernel and the projection kernel. We can summarize the new MLEM approach in matrix notation, based on the description from the GPU-PET reconstructor 3.6,

$$X^{(n+1)} = \frac{X^{(n)}}{\Lambda} G * \left(P^T \frac{Y}{\tilde{Y}_{MC}^{(n)}} \right), \tag{6.1}$$

where $X^{(n)}$ is the image at iteration n, G the PSF image blurring kernel (see the details in section 3.2.1.1), P^T the backprojection kernel (see section 3.2.1.2), $\tilde{Y}_{MC}^{(n)}$ is the array of expected projections at iteration n for the MC projection method, Y the array of detected events, and Λ is the sensitivity image. The sensitivity image is computed in advance, based on the backprojection scheme and the normalization factors for the UMC-PET projector,

$$\begin{aligned} \Lambda &= G * \left(P^T V_{MC} \right) \\ v_{i,MC} &= \frac{q_i}{\tilde{q}_{i,MC}} \frac{\tilde{l}_i}{l_i}, \end{aligned} \tag{6.2}$$

where V_{MC} is the normalization for the MC projection method, that is element-wise detailed in the second line ($v_{i,MC}$). q_i is the measured calibration obtained from a cylindrical acquisition, $\tilde{q}_{i,MC}$ is the equivalent LOR projection based on a simulation of a cylindrical source filling the whole FOV, and $\tilde{l}_i/l_i$ is the LOR length ratio in case the calibration cylinder and the simulated FOV do not have equal size. Note that this normalization approach is similar to the one proposed in equation 3.4 of section 3.2.1.3.

The MLEM algorithm [Shepp and Vardi, 1982] was chosen instead of the OSEM [Hudson and Larkin, 1994] because the list of simulated events voxel-LOR has a random distribution, and we need to read the whole list to produce a projection estimation (the backprojection time is negligible in this approach). We did not use regularization techniques in the reconstruction since we wanted to compare the SRM performance in the case of accurate MC modeling, and further MC modeling in the projection step already warrants a large amount of regularization. We also used a low-noise input image as X^0, based on the GPU-PET reconstructed one with strong regularization (β=1 and median root prior, see section 3.2.2.1) to speed-up the convergence of this method. It is convenient to comment that for this method to be practical, the full list of voxel-LOR events has to fit within the RAM memory of the reconstruction computer. Nowadays it is common to have server computers with >512 GB of RAM memory for under 3000 euros, such as the one employed for these computations, thus this limitation is not of large concern.

We have implemented two projection modes to estimate $\tilde{Y}_{MC}$ with the UMC-PET simulator, based on a fully-MC projection (UMC-PET projection), and a MC based projection decomposition using the system symmetries (symUMC-PET projection).

6.2.2.1 UMC-PET projection

We have included the whole SRM in an MC simulation. The µ-map is inserted inside a model of the scanner, and an homogeneous source filling the region of the FOV where we are interested is simulated with the UMC-PET simulator. All the effects of emission, scatter and attenuation through the patient body, and detector response are accounted for. The simulation generates a list of voxel-LOR detected coincidences (8 bytes per event, 2 integers of 4 bytes), that will be used for the forward projection step. To reduce the list size, we only keep LORs that have events, as null LORs do not contribute to the image update. To estimate the projection, we run over the whole list accumulating the value of voxel x_{j_l} in the coincidence LOR $\tilde{y}_{i,MC}$ for every event in the list l,

$$\tilde{y}_{i,MC} = \sum_l x_{j_l} \delta(i_l, i), \tag{6.3}$$

where δ is the Kronecker delta.

Since all the voxels have equal emission probability in the list, the value of $\tilde{y}_{i,MC}$ is proportional to the detection probability in the LOR and the source activity (we can identify the SRM elements as $a_{ij} = \sum_l \delta(j_l, j)\delta(i_l, i)$), providing the expected activity. The UMC-PET simulator does not estimate random events whose contribution will have to be accounted separately. In this study we subtracted the random events contribution from the data.

6.2.2.2 UMC-PET projection decomposition with symmetries

In this case, an homogeneous cylinder centered in the scanner without the µ-map was simulated. Therefore, we do not account for the scatter and the attenuation in the simulation, which allows us to increase the statistics in the projection using the symmetries of the scanner (we call this projection method symUMC-PET, see figure 6.1). To avoid aliasing, the cartesian coordinates of the emission voxel and the LOR index were recorded in the list of events generated (10 bytes per event, 1 integer of 4 bytes, and 3 short 2 bytes integers encoding the emission position), and further empty LORs

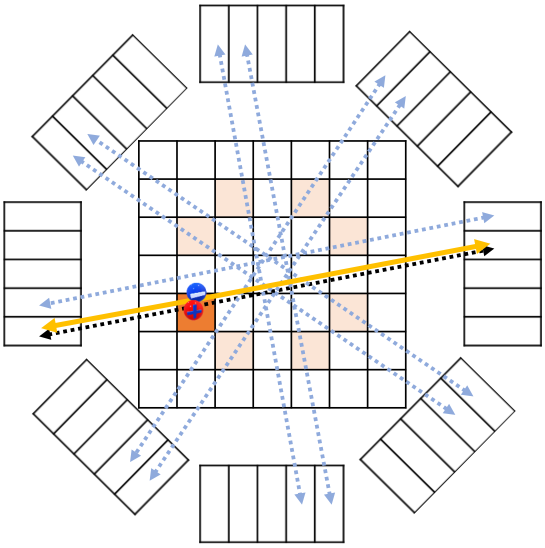


Figure 6.1: Schematic representation of the rotational symmetries in a cylindrical scanner of 8 blocks with 5 crystals per block. The dashed black line represents the original LOR of a simulated event, and the dashed light blue lines represent all the symmetric LORs. The shadowed squares represent the equivalent emission voxels for every symmetric LOR.

must be saved as well, unless all the symmetric LORs have no counts. The summation over the list of events reads:

$$\tilde{y}_{i,MC} = s_i + u_i \cdot \sum_l x(l,i)\delta(k_l,k_i), \tag{6.4}$$

where s_i and u_i are scatter and attenuation coefficients, k_i and k_l are the symmetry indices of LORs i and the LOR l in the list, and $x(l,i)$ is the value of the emission voxel for the symmetric LOR i given the detection in LOR l.

For every event in the list we identify its symmetry index and calculate the relative position with the LOR. Then we run a loop through a list of all the symmetric LORs, and we sum all the equivalent emission voxels in each of them (figure 6.1). Notice that in this implementation the more symmetries are accounted for, the larger is the computational cost. For this reason we only accounted for block rotational symmetries in this study, we did consider border effects inside a block.

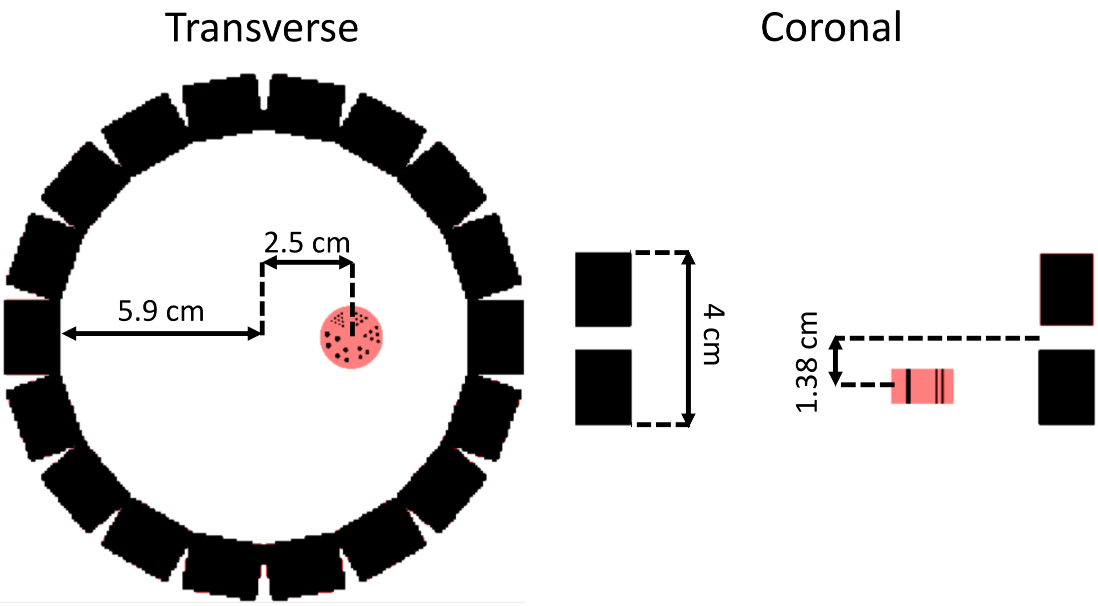


Figure 6.2: Schematic transverse and coronal view of the 2R-Argus scanner with the Derenzo phantom.

6.2.3 PET data acquisitions

6.2.3.1 Simulated Derenzo phantom in the 2R-Argus scanner

We have simulated a Derenzo phantom in the 2R-Argus scanner to show visual image improvement when the MC projection method is used compared with the GPU-PET reconstructor. The details of the Argus scanner can be found in section 2.2.1, and the Derenzo phantom was described in section 2.5.1 (right-hand side of figure 2.8). We shifted the Derenzo 25 mm off center in the transverse direction and 13.8 mm in the axial direction (see figure 6.2) to explore the parallax error in the acquisition.

We ran the simulation with the UMC-PET simulator (see chapter 5). We simulated $1{\cdot}10^{11}$ decays of ^{18}F-FDG, equivalent to an acquisition of 60 minutes of 206 MBq (5.57 mCi). This value is higher than the usual activity values used in PET since we wanted to test the reconstruction in optimal scenarios of low statistical noise. The total detected events were $1.78{\cdot}10^{9}$. We did not include an attenuation object in the simulation. An homogeneous cylinder of 4 cm length and 7 cm diameter was used as calibration acquisition for the reconstruction.

6.2.3.2 Image Quality phantom in the 6R-SuperArgus scanner

We show the data acquired with an Image Quality phantom from the NEMA NU4-2008 [National Electrical Manufacturers Association, 2008] (the details of the phantom might be found in 2.3) in the 6R-SuperArgus scanner (see section 2.2.1). Quantitative measurements of the noise, recovery coefficients (RC), and the spill over ratio (SOR) have been evaluated as explained in the section 2.3. The phantom was filled with a solution of 200 µCi of ^{18}F, acquired for 20 minutes ($1.08 \cdot 10^8$ coincidences) with an energy window of 425-600 keV. In comparison with the official NEMA NU4-2008 protocol, we have used twice the activity (100 µCi were proposed in the protocol).

6.3 Results

6.3.1 Simulated Derenzo phantom in the 2R-Argus scanner

In figure 6.3 we show the image reconstructions from the simulated Derenzo phantom in the 2R-Argus scanner. The parallax error improvement when the UMC-PET projection is used can be noticed in the images and in the line profiles. The GPU-PET reconstruction used a PSF of 1.0 mm in the transverse direction and 1.4 mm in the axial direction. In the case of SV-PSF, PSF gradually increases in the radial and axial directions. The voxel size used was of $0.2 \times 0.2 \times 0.782$ mm^3.

The list of events voxel-LOR simulated for the UMC-PET projection was ran with an homogeneous cylindrical source of 2 cm diameter and 1 cm length centered at the position of the original phantom (approximately $1.00 \cdot 10^5$ voxels, representing a 2% of the total FOV). The simulation ran $1 \cdot 10^{11}$ decays in 17.8 minutes, from which we obtained $1.75 \cdot 10^9$ coincidences (13 Gb). The number of active LORs in this acquisition was $4.45 \cdot 10^6$. This resulted in an average of $3.93 \cdot 10^2$ coincidences per LOR. The reconstruction required 3.31 seconds per iteration, resulting in 10 minutes to achieve 180 iterations. Both simulation and reconstruction were performed in an Intel(R) Xeon(R) W-2155 CPU @ 3.30GHz for the host, and a 11 GB GeForce RTX 2080 Ti GPU, with 4352 cores, for the GPU.

We did not implement the symUMC-PET projection here because the source was located in a reduced region of the FOV with minor gain from the symmetries.

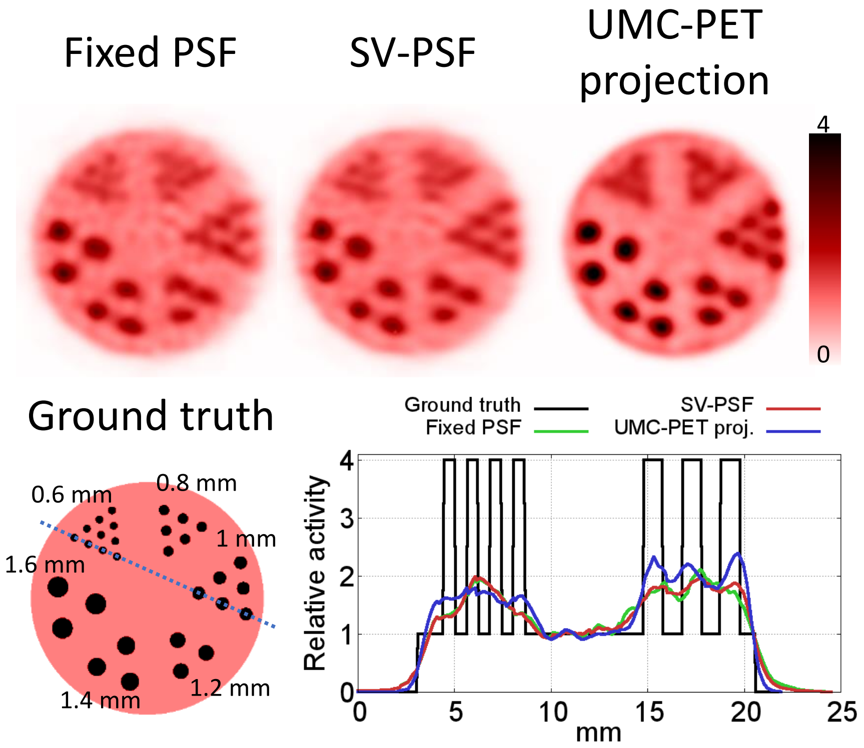


Figure 6.3: Top row, from left to right, the GPU-PET standard reconstruction with fixed PSF (Fixed PSF) and with spatially variant PSF (SV-PSF), and the proposed approach with the UMC-PET projector. Transverse view of 5 mm collapsed slices are shown for each reconstruction. A scheme of the ground truth phantom with a line profile through the 0.6 mm and 1 mm rods is shown in the bottom row. All the images have 180 MLEM updates. The noise measured in a cylindrical ROI in the center of the phantom is 13.5% (Fixed PSF), 13.6% (SV-PSF) and 11.2% (UMC-PET projector). The rod to background activity ratio is 4:1.

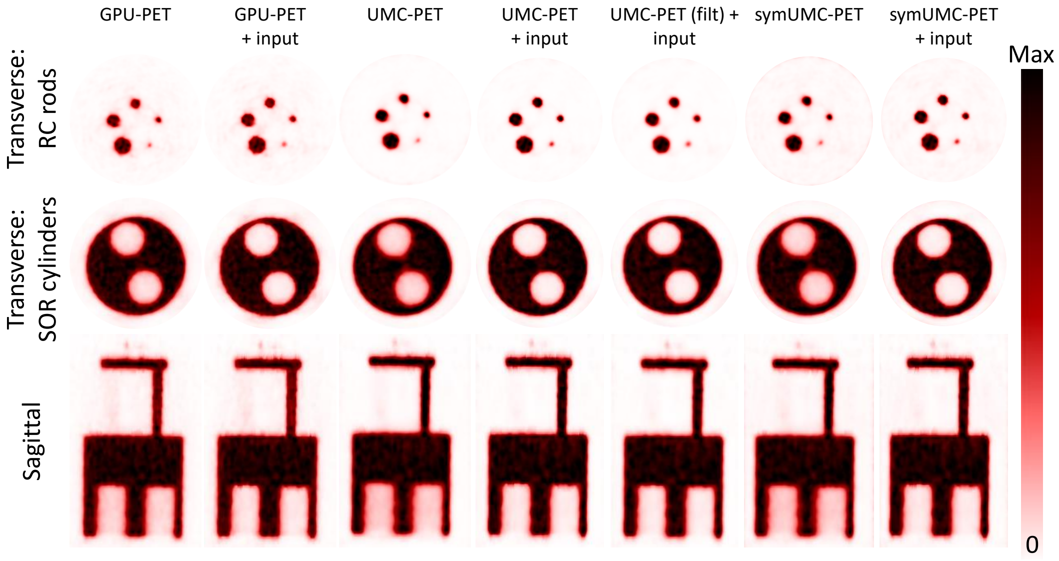


Figure 6.4: Transverse view of one single slice of the RC rods and the SOR cylinders, and sagittal view of the IQ phantom for the reconstruction methods proposed. The images have an equivalent noise level around 5% (details in table 6.1). color scale has been set to 1.1 times the average activity in the uniform region.

6.3.2 Image Quality phantom in the 6R-SuperArgus scanner

In figure 6.4 we show the images obtained with the methods described to reconstruct a real IQ phantom acquisition in the 6R-SuperArgus scanner. In this figure we compare the standard GPU-PET reconstructions with the UMC-PET projection method and the symUMC-PET projection. Both MC projection methods included the case with an input image, and for the UMC-PET projection we also added a Gaussian image post-filtering of 0.75 mm FWHM to reduce noise. All images have a voxel size of $0.25\times0.25\times0.81$ mm^3, and the PSF implemented was of 1.0 mm in the transverse direction and 1.4 mm in the axial direction. The SV-PSF method was not implemented since the phantom lies in the central region of the FOV where the PSF is homogeneous. The quantitative measurements of noise, RC, and SOR are presented in table 6.1. In figures 6.5 and 6.6 we show the evolution of noise-RC and noise-SOR for the methods proposed, with different number of iterations.

All the results from this section were obtained on an Intel(R) Xeon(R) W-2155 CPU @ 3.30GHz CPU host unit, and a 23.7 GB GeForce RTX 3090 with 10496 cores GPU device.

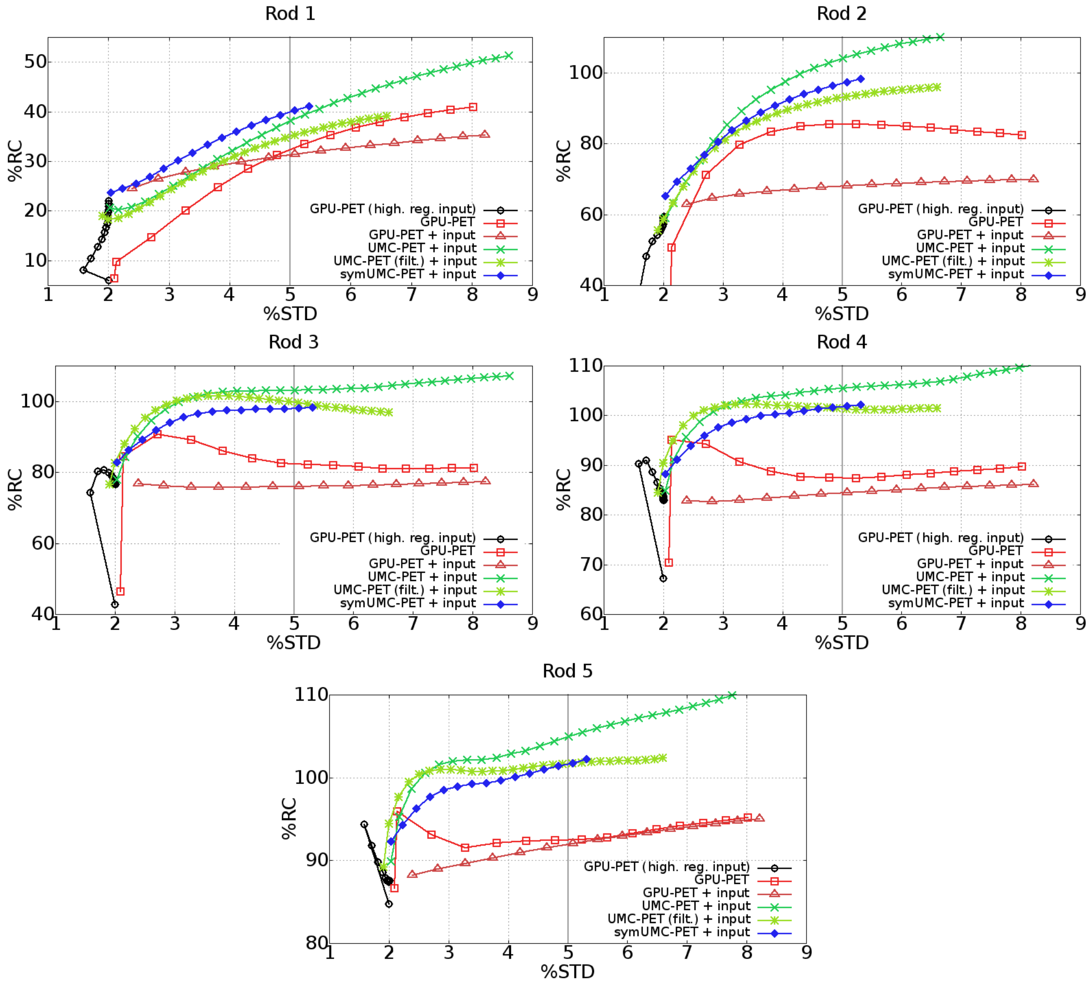


Figure 6.5: RC-noise curves for the 5 rods in the IQ phantom. The horizontal gray line shows the 5% noise level. We show the progress per every iteration of 6 subsets for the GPU-PET reconstructor, and every 2 iterations in the MC projection methods.

Noise, RC, and SOR measured values								
-	%STD	%RC					%SOR	
		1 mm	2 mm	3 mm	4 mm	5 mm	Water	Air
GPU-PET	5.23	33.5	85.5	82.2	87.3	92.6	5.86	6.18
GPU-PET + input	5.08	31.4	68.0	76.0	84.5	92.0	3.49	3.45
UMC-PET	5.10	25.0	120.9	122.7	114.2	109.6	7.00	7.96
UMC-PET + input	5.01	38.1	104.6	103.0	105.5	105.0	2.85	3.11
UMC-PET (filt) + input	5.08	35.2	93.2	99.5	101.3	101.7	2.53	2.82
symUMC-PET	5.06	28.4	114.3	117.0	110.7	106.5	7.30	7.89
symUMC-PET + input	5.08	40.2	97.3	98.0	101.9	101.7	3.00	3.04

Table 6.1: Noise, RC, and SOR for the IQ phantom at an equivalent noise level in the 6R-Super Argus scanner for all the images in figure 6.4.

The simulation for the UMC-PET projection was performed with an homogeneous cylinder of 4 cm diameter and 8 cm length (a total of $1.99{\cdot}10^6$ voxels, a 6% of the total FOV). The IQ µ-map was introduced in the simulation to account for scatter and attenuation in the projection. We ran $5{\cdot}10^{12}$ decays in the energy window of 425-600 keV, that resulted in a list of $6.45{\cdot}10^{10}$ coincidences (481 Gb) in 12 hours. The active LORs in this case were of $5.53{\cdot}10^7$, thus obtaining $1.17{\cdot}10^3$ events per LOR. The reconstruction time per iteration was of 4 minutes and 37 seconds, thus resulting in 4 hours and 37 minutes for 60 iterations. The total reconstruction (simulation plus reconstruction) took ∼17 hours.

In the symUMC-PET case, we simulated the same source without the µ-map. We ran $1.6{\cdot}10^{12}$ decays in the energy window of 425-600 keV. The events list in this case required a similar amount of memory, having $5.15{\cdot}10^{10}$ coincidences (480 Gb; recall that the list for the symmetries needed 10 bytes per event, instead of 8 bytes). The time per iteration increased to 14 minutes and 12 seconds, resulting in 14 hours to achieve 60 iterations. In this case, the simulated list of events can be precomputed to save time, since it is not object-dependent. This example included only rotational symmetries.

6.4 Discussion

The implementation of the UMC-PET projection is mainly limited by the computation of the projection step and the list of simulated events. In the case of the simulated Derenzo from section 6.3.1 we were able to achieve the total reconstruction in less than

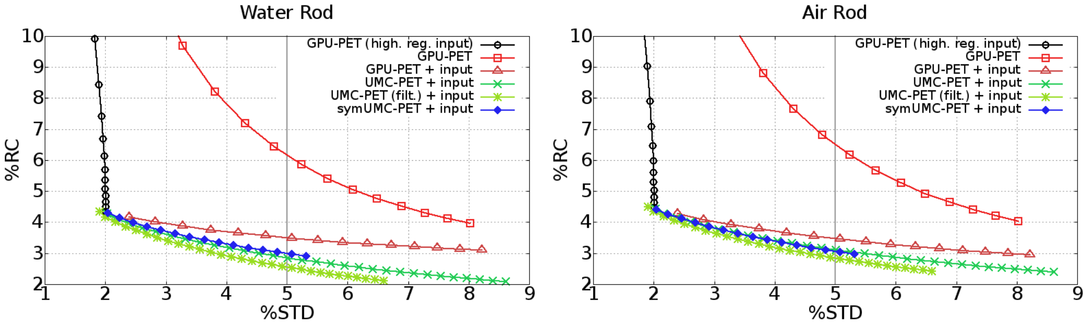


Figure 6.6: SOR-noise curves for the water and air cylinders in the IQ phantom. The horizontal gray line shows the 5% noise level. We show the progress per every iteration of 6 subsets for the GPU-PET reconstructor, and every 2 iterations in the MC projection methods.

30 minutes (17.8 minutes for the simulation and 10 minutes for the image reconstruction). For the IQ phantom in the 6R-SuperArgus scanner from section 6.3.2, the number of voxels needed to cover the phantom was ∼20 times higher and the number of active LORs was ∼12 times larger, increasing the simulation time needed to achieve reasonable statistics up to 12 hours, and the reconstruction time to more than 4 hours. Furthermore, we reduced the simulation and reconstruction to a small space of the FOV to optimize the statistics in both cases. For these reasons. While the feasibility of a fully-MC projection scheme for image reconstruction in large scanners for routine use is not achieved, due to the large reconstruction times, MC reconstructions can be employed as a gold standard for specific cases, for instance to guide refinements of the more standard reconstruction and to identify the contributions to the SRM which are more relevant for image quality. However, it is worth noting that full MC reconstruction method can already be applied routinely for smaller scanners, such as the 2R-Argus and the Super-Argus Compact or 3R-SuperArgus, where reconstruction times will be below one hour, and that larger scanner will soon amenable to these methods, taking into account the exponential grow of RAM memory amount in PCs and computing capabilities of GPUs.

It can be argued that the OSEM [Hudson and Larkin, 1994] algorithm might be used to reduce the reconstruction time. However, to efficiently implement OSEM here, we would need to process the list first, grouping events involved in the same LOR. The size of the events list and the number of LORs involved prevented us from attempting this approach, but it is open for further study.

A second limitation of the proposed method is the size of the list of voxel-LOR events. In the results shown in section 6.3.2 for the 6R-SuperArgus scanner, the total size of this list was 481 Gb. The system used for the computational calculations had a RAM memory of 512 GB, which is becoming more and more available in server PCs. The sparsity and symmetries implemented in [Chien-Min Kao et al., 2007, Meng et al., 2019, Rafecas et al., 2004, Zhang et al., 2010] to reduce storage needs cannot be exploited because of scatter and attenuation modeling.

The pure MC projection (UMC-PET projection) includes statistical MC noise that is transferred to the image during reconstruction. We introduced a very low-noise initial image to allow for faster convergence and image post-filtering to reduce overshooting, noticeably improving the performance of the RC-noise and SOR-noise measurements (table 6.1 and figures 6.5 and 6.6). When no initial image was employed, the 5% noise level is reached with a few iterations, before the convergence of the SOR and the RC in the smallest rod is achieved. The rods from 2 to 5 mm converge in the first iterations, but due to the noise, overshoot is seen. We found that larger statistics in the UMC-PET projection reduced the overshoot in these rods whilst increasing the RC in the smaller rod of 1 mm diameter. In the case of the symUMC-PET projection, the higher statistics provided by the symmetries shows an example of this behaviour, but in this case we had to separate the attenuation and scatter components, slightly degrading the SOR-noise ratio.

Mostly for sake of completeness, we would like to use the UMC-PET coincidence list of events to perform the backprojection as well. This would allow modeling scatter in the backprojection, completing the MC forward/backward projection model. However, we have verified (as expected) that the statistics achieved from our studies resulted in very noisy images when the UMC-PET was introduced also in the back-projection.

Even though the statistics implies a limitation for the UMC-PET projection, we can see appreciable improvements in the quality of the images obtained with this method. In figure 6.3, we showed very good ability to correct for parallax error, outperforming the spatially-variant PSF. In table 6.1, the best RC values were obtained with the symUMC-PET projection, with input image. The best SOR values are obtained for the UMC-PET projection with post-fitering. In general both UMC-PET-based methods (with or without symmetries) outperformed the traditional scheme applied with the GPU-PET reconstructor from chapter 3.

On the one hand, from these results we can conclude that the MC model improves the image resolution and signal-to-noise ratio. Many previous works have explored MC reconstructions [Chien-Min Kao et al., 2007, Herraiz et al., 2006a, Meng et al., 2019, Pilleri et al., 2019, Rafecas et al., 2004, Southekal et al., 2011, Wei and Vaska, 2020, Xu et al., 2019, Zhang et al., 2010, Zhou and Qi, 2014]. On the other hand, the implementation of MC-based attenuation and scatter led us to the best SOR values. In the case of scatter, the UMC-PET projector computes a new scatter estimation after every image update, and this is one of the main soruces of noise in the reconstruction.AI enhanced scatter estimations might be a practical solution to reduce the time required to obtain noise-reduced UMC-PET scatter estimates.

Two compact choices to store the SRM are the use of TORs [Herraiz et al., 2006a, Markiewicz et al., 2014] or PSF in the projection domain [Gong et al., 2017, Tohme and Qi, 2009]. Compared against image PSF, these methods allow to model the differences in the LORs that overlap in the image space, for example, because of the border effects in the corners of a pixellated detector, that are averaged if we model the response in image space. Usually the MC simulation of a complete SRM requires large computational times. Pseudo-symmetries that neglect border effects are implemented to reduce this effort. [Chien-Min Kao et al., 2007, Herraiz et al., 2006a, Panin et al., 2006, Zhou and Qi, 2011]. Thanks to the UMC-PET simulation speed, we can generate larger number of events in shorter times, avoiding these approximations.

Chapter 7

Conclusions of this thesis

In this thesis, we studied if new simulation and image reconstruction methods based on GPU computing and realistic physics modeling could be able to achieve improved image quality in PET with respect to the current state-of-the-art.

- **GPU-based PET image reconstruction software**

  We have developed the GPU-PET reconstructor, a flexible tool for image reconstruction that successfully adapted to different data formats (sinogram, LOR histograms, and list-mode data) and scanner geometries (cylindrical and spherical), allowing fully 3D reconstructions with up to a billion LORs within a few minutes. The traditional MAP-OSEM algorithm with image blurring PSF for resolution modeling and two different Bayesian priors allowed to achieve submilimeter resolution in the real acquisitions of the 6R-SuperArgus scanner.

- **Improved Image Reconstruction with Super-iterations**

  A novel algorithm for data-driven improved sampling in PET imaging has demonstrated to achieve noticeable image quality gain of ∼10% in the recovery coefficients of the image quality phantom of the NEMA NU4-2008, acquired with the 4R-SuperArgus scanner, without raising the noise level. This is something remarkable if we take into account that this is obtained respect to an already optimized algorithm. The algorithm was implemented in sinogram data from a simplified simulation, and in the GPU-PET reconstructor with the MAP-OSEM algorithm, proving its compatibility with any data modality and reconstruction algorithm.

The usage of GPUs allows the reconstruction time to be kept within reasonable limits.

- **Ultra-fast Monte Carlo PET simulator**

  We have shown that the UMC-PET simulator is a fast, accurate and flexible Monte Carlo code for PET simulation accelerated with GPU, developed to focus on reconstruction assessment (scatter correction, or SRM estimation) and scanner design. The precomputation of the relevant physical quantities, and a factorized scheme of the main physics underlying the technique simplified the code, and they did not negatively affect the simulator accuracy compared against PeneloPET. We have proven that the novel definition of voxelized scanner and detectors makes it easy its use for any geometry and most of the currently used detector configurations. We have given a proof of practical use for MC scatter estimation using the UMC-PET, outperforming the SSS algorithm in a reasonable time in the high contrast regions where the quantification is more challenging. Though the UMC-PET simulator will be freely available for most purposes, it is not open yet since preparation of user friendly documentation is still pending.

- **Full integrated tool: UMC-PET improved reconstruction**

  We have shown that is is possible to use a fully MC-projection based on the speed-ups of the UMC-PET simulator (including non-collinearity, DOI effects, scatter, attenuation and the detectors response) combined with the GPU-PET reconstructor, which outperformed the traditional methods in terms of visual image quality, RC-noise, and SOR-noise. Variance reduction methods, such as the use of symmetries, and image post-filtering proved to reduce the effect of limited statistics in the MC-projection. The computational burden of the method did not allow for daily use, but it can be used as a gold reference for current particular cases and, in the near future, with scanners with more computing capabilities. In any case, the computing power of GPUs is expected to continue increasing in the following years, and therefore, the method proposed in this thesis will become more feasible for routine practice.

As a general conclusion, the aid of GPU parallel computing allowed us to develop tools and methods in PET reconstruction and simulation, paving the way for next generation PET scanners and driving image reconstruction to no-compromised optimal performance.

Thanks to the flexibility and power of both the UMC-PET simulation and the GPU-PET reconstruction packages, these tools have been combined several times to assess the image quality of the scanners of the SuperArgus family, and a previous version of the GPU-PET reconstructor is currently being used in the commercial scanners. In the case of the prototype HSTR-BrainPET with spherical geometry, our assessment of the scanner performance ended up as part of a grant proposal to the National Institute of Health (NIH) of USA, which was eventually funded and is currently ongoing [Catana, 2019a].

Chapter 8

Acknowledgements

We thank Juan José Vaquero López (Universidad Carlos III, Madrid, Spain) and Manuel Desco (Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain) for invaluable support during the acquisitions of the 6R-SuperArgus and the 4R-SuperArgus scanners.

We thank Ciprian Catana (Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts, United States of America) for providing the detailed geometry of the HSTR-BrainPET scanner.

We thank Onofrio Catalano (Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts, United States of America) for providing the data from the Biograph mMR scanner.

We acknowledge Fernando Arias Valcayo, who adapted the GPU-PET reconstructor to use spatially variant PSF.

List of Figures

List of Tables

References

Abbaszadeh, S., Chinn, G., and Levin, C. S. (2018). Positioning true coincidences that undergo inter-and intra-crystal scatter for a sub-mm resolution cadmium zinc telluride-based PET system. *Physics in Medicine and Biology*, 63(2). 87, 107

Abushab, K. M., Herraiz, J. L., Vicente, E., Cal-Gonzalez, J., Espana, S., Vaquero, J. J., Jakoby, B. W., and Udias, J. M. (2016). Evaluation of PeneloPET Simulations of Biograph PET/CT Scanners. *IEEE Transactions on Nuclear Science*, 63(3):1367–1374. 49, 108

Adam, L.-E., Karp, J. S., and Brix, G. (1999). Investigation of scattered radiation in 3D whole-body positron emission tomography using Monte Carlo simulations. *Physics in Medicine and Biology*, 44(12):2879–2895. 32, 107, 136

Afshar-Oromieh, A., Wolf, M., Haberkorn, U., Kachelrieß, M., Gnirs, R., Kopka, K., Schlemmer, H.-P., and Freitag, M. T. (2017). Effects of arm truncation on the appearance of the halo artifact in 68Ga-PSMA-11 (HBED-CC) PET/MRI. *European Journal of Nuclear Medicine and Molecular Imaging*, 44(10):1636–1646. 32, 136, 137, 146

Agency, N. E. (2019). *PENELOPE 2018: A code system for Monte Carlo simulation of electron and photon transport.* 6, 7, 37, 39, 40, 41, 42, 43, 44, 108, 110, 116, 169, 170

Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., Barrand, G., Behner, F., Bellagamba, L., Boudreau, J., Broglia, L., Brunengo, A., Burkhardt, H., Chauvie, S., Chuma, J., Chytracek, R., Cooperman, G., Cosmo, G., Degtyarenko, P., Dell'Acqua, A., Depaola, G., Dietrich, D., Enami, R., Feliciello, A., Ferguson, C., Fesefeldt, H., Folger, G., Foppiano, F., Forti, A., Garelli, S., Giani, S., Giannitrapani, R., Gibin, D., Gomez Cadenas, J. J., Gonzalez, I., Gracia Abril, G., Greeniaus, G., Greiner, W., Grichine, V., Grossheim, A., Guatelli, S., Gumplinger, P., Hamatsu, R., Hashimoto, K., Hasui, H., Heikkinen, A., Howard, A., Ivanchenko, V., Johnson, A., Jones, F. W., Kallenbach, J., Kanaya, N., Kawabata, M., Kawabata, Y., Kawaguti, M., Kelner, S., Kent, P., Kimura, A., Kodama, T., Kokoulin, R., Kossov, M., Kurashige, H., Lamanna, E., Lampen, T., Lara, V., Lefebure, V., Lei, F., Liendl, M., Lockman, W., Longo, F., Magni, S., Maire, M., Medernach, E., Minamimoto, K., Mora de Freitas, P., Morita, Y., Murakami, K., Nagamatu, M., Nartallo, R., Nieminen, P., Nishimura, T., Ohtsubo, K., Okamura, M., O'Neale, S., Oohata, Y., Paech, K., Perl, J., Pfeiffer, A., Pia, M. G., Ranjard, F., Rybin, A., Sadilov, S., di Salvo, E., Santin, G., Sasaki, T., Savvas, N., Sawada, Y., Scherer, S., Sei, S., Sirotenko, V., Smith, D., Starkov, N., Stoecker, H., Sulkimo, J., Takahata, M., Tanaka, S., Tcherniaev, E., Safai Tehrani, E., Tropeano, M., Truscott, P., Uno, H., Urban, L., Urban, P., Verderi, M., Walkden, A., Wander, W., Weber, H., Wellisch, J. P., Wenaus, T., Williams, D. C., Wright, D., Yamada, T., Yoshida, H., and Zschiesche, D. (2003). GEANT4 - A simulation toolkit. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303. 108

Ahn, S. and Fessler, J. (2003). Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Transactions on Medical Imaging*, 22(5):613–626. xviii, 23, 27, 28

Ahn, S., Ross, S. G., Asma, E., Miao, J., Jin, X., Cheng, L., Wollenweber, S. D., and Manjeshwar, R. M. (2015). Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. *Physics in Medicine and Biology*, 60(15):5733–5751. xviii, 28, 69, 84

Akamatsu, G., Tashima, H., Takyu, S., Kang, H. G., Iwao, Y., Takahashi, M., Yoshida, E., and Yamaya, T. (2021). Design consideration of compact cardiac TOF-PET systems: a simulation study. *Physics in Medicine and Biology*, 66(7):074002. 144

Aklan, B., Jakoby, B. W., Watson, C. C., Braun, H., Ritt, P., and Quick, H. H. (2015). GATE Monte Carlo simulations for variations of an integrated PET/MR hybrid imaging system based on the Biograph mMR model. *Physics in Medicine and Biology*, 60(12):4731–4752. 107

Alenius, S. and Ruotsalainen, U. (1997). Bayesian image reconstruction for emission tomography based on median root prior. *European Journal of Nuclear Medicine and Molecular Imaging*, 24(3):258–265. xviii, xix, xxi, xxvi, 23, 28, 62, 68, 84

Alenius, S., Ruotsalainen, U., and Astola, J. (1999). Attenuation correction for PET using count-limited transmission images reconstructed with median root prior. *IEEE Transactions on Nuclear Science*, 46(3):646–651. 28, 31

Alerstam, E., Andersson-Engels, S., and Svensson, T. (2008a). White Monte Carlo for time-resolved photon migration. *Journal of Biomedical Optics*, 13(4):041304. 109

Alerstam, E., Svensson, T., and Andersson-Engels, S. (2008b). Parallel computing with graphics processing units for high-speed Monte Carlo simulation of photon migration. *Journal of Biomedical Optics*, 13(6):060504. 44, 109

Alessio, A. and Kinahan, P. (2006). PET Image Reconstruction. In *Nuclear Medicine*. Elsevier. 20, 21, 27

Alessio, A., Kinahan, P., Harrison, R., and Lewellen, T. (2005). Measured Spatially Variant System Response for PET Image Reconstruction. In *IEEE Nuclear Science Symposium Conference Record, 2005*, volume 4, pages 1986–1990. IEEE. 25, 83

Allison, J., Amako, K., Apostolakis, J., Arce, P., Asai, M., Aso, T., Bagli, E., Bagulya, A., Banerjee, S., Barrand, G., Beck, B., Bogdanov, A., Brandt, D., Brown, J., Burkhardt, H., Canal, P., Cano-Ott, D., Chauvie, S., Cho, K., Cirrone, G., Cooperman, G., Cortés-Giraldo, M., Cosmo, G., Cuttone, G., Depaola, G., Desorgher, L., Dong, X., Dotti, A., Elvira, V., Folger, G., Francis, Z., Galoyan, A., Garnier, L., Gayer, M., Genser, K., Grichine, V., Guatelli, S., Guèye, P., Gumplinger, P., Howard, A., Hřivnáčová, I., Hwang, S., Incerti, S., Ivanchenko, A., Ivanchenko, V., Jones, F., Jun, S., Kaitaniemi, P., Karakatsanis, N., Karamitros, M., Kelsey, M., Kimura, A., Koi, T., Kurashige, H., Lechner, A., Lee, S., Longo, F., Maire, M., Mancusi, D., Mantero, A., Mendoza, E., Morgan, B., Murakami, K., Nikitina, T., Pandola, L., Paprocki, P., Perl, J., Petrović, I., Pia, M., Pokorski, W., Quesada, J., Raine, M., Reis, M., Ribon, A., Ristić Fira, A., Romano, F., Russo, G., Santin, G., Sasaki, T., Sawkey, D., Shin, J., Strakovsky, I., Taborda, A., Tanaka, S., Tomé, B., Toshito, T., Tran, H., Truscott, P., Urban, L., Uzhinsky, V., Verbeke, J., Verderi, M., Wendt, B., Wenzel, H., Wright, D., Wright, D., Yamashita, T., Yarba, J., and Yoshida, H. (2016). Recent developments in Geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225. 108

Alva-Sánchez, H., Quintana-Bautista, C., Martínez-Dávalos, A., Ávila-Rodríguez, M. A., and Rodríguez-Villafuerte, M. (2016). Positron range in tissue-equivalent materials: experimental microPET studies. *Physics in Medicine and Biology*, 61111(17):6307–6321. 4, 118

Álvarez-Gómez, J. M., Santos-Blasco, J., Moliner Martínez, L., and Rodríguez-Álvarez, M. J. (2021). Fast Energy Dependent Scatter Correction for List-Mode PET Data. *Journal of Imaging*, 7(10):199. 16, 33

Andersen, A. (1984). Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94. 22

Anger, H. O. (1969). *Scintillation Camera and Multiplane Tomographic Scanner.* Lawrence Berkeley National Laboratory, lbnl repor edition. 13

Arabi, H., AkhavanAllaf, A., Sanaat, A., Shiri, I., and Zaidi, H. (2021). The promise of artificial intelligence and deep learning in PET and SPECT imaging. *Physica Medica*, 83(February):122–137. xviii, 18, 36

Arabi, H., Bortolin, K., Ginovart, N., Garibotto, V., and Zaidi, H. (2020). Deep learning-guided joint attenuation and scatter correction in multitracer neuroimaging studies. *Human Brain Mapping*, 41(13):3667–3679. 33, 36, 137

Arabi, H. and Zaidi, H. (2021). Assessment of deep learning-based PET attenuation correction frameworks in the sinogram domain. *Physics in Medicine & Biology*, 66(14):145001. 36

REFERENCES

Arias-Valcayo, F., Herraiz, J. L., Galve, P., Vaquero, J. J., Desco, M., and Udias, J. M. (2019). Awake preclinical brain PET imaging based on point sources. In Matej, S. and Metzler, S. D., editors, *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, number May, page 67. SPIE. 16, 74, 82

Arias-Valcayo, F., Herraiz, J. L., Lopez-Montes, A., Vaquero, J. J., Desco, M., Galve, P., and Udias, J. M. (2020). No gates , Real-Time Cardiac Reconstruction of Small Animal PET Data. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. 16, 82

Asma, E., Ahn, S., Ross, S. G., Chen, A., and Manjeshwar, R. M. (2012). Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, number x, pages 4062–4066. IEEE. 28

Augusto, R., Bauer, J., Bouhali, O., Cuccagna, C., Gianoli, C., Kozłowska, W., Ortega, P., Tessonnier, T., Toufique, Y., Vlachoudis, V., Parodi, K., and Ferrari, A. (2018). An overview of recent developments in FLUKA PET tools. *Physica Medica*, 54(June):189–199. 108

Badal, A. and Badano, A. (2009). Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Medical Physics*, 36(11):4878–4880. 44, 109, 110, 123

Badawi, R., Kohlmyer, S., Harrison, R., Vannoy, S., and Lewellen, T. (2000). The effect of camera geometry on singles flux, scatter fraction and trues and randoms sensitivity for cylindrical 3D PET-a simulation study. *IEEE Transactions on Nuclear Science*, 47(3):1228–1232. 34

Badawi, R. D., Lodge, M. A., and Marsden, P. K. (1998). Algorithms for calculating detector efficiency normalization coefficients for true coincidences in 3D PET. *Physics in Medicine and Biology*, 43(1):189–205. 30

Badawi, R. D. and Marsden, P. K. (1999). Developments in component-based normalization for 3D PET. *Physics in Medicine and Biology*, 44(2):571–594. 30, 35, 66

Badawi, R. D., Miller, M. P., Bailey, D. L., and Marsden, P. K. (1999). Randoms variance reduction in 3D PET. *Physics in Medicine and Biology*, 44(4):941–954. 34

Bailey, D. L., Townsend, D. W., Valk, P. E., and Maisey, M. N., editors (2005). *Positron Emission Tomography: Basic Sciences*. Springer-Verlag, London. 9, 10, 29, 30, 31, 33, 179

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800. 36

Barker, W., Thada, S., and Dieckmann, W. (2009). A GPU-accelerated implementation of the MOLAR PET reconstruction package. In *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 4114–4119. IEEE. 23, 61, 83

Baró, J., Sempau, J., Fernández-Varea, J., and Salvat, F. (1995). PENELOPE: An algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 100(1):31–46. 108, 116

Barret, O., Carpenter, T. A., Clark, J. C., Ansorge, R. E., and Fryer, T. D. (2005). Monte Carlo simulation and scatter correction of the GE Advance PET scanner with SimSET and Geant4. *Physics in Medicine and Biology*, 50(20):4823–4840. 108

Bauert, J., Tessonnier, T., Debus, J., and Parodi, K. (2019). Offline imaging of positron emitters induced by therapeutic helium, carbon and oxygen ion beams with a full-ring PET/CT scanner: experiments in reference targets. *Physics in Medicine & Biology*, 64(22):225016. 107

Baum, K. G. and Helguera, M. (2007). Execution of the SimSET Monte Carlo PET/SPECT Simulator in the Condor Distributed Computing Environment. *Journal of Digital Imaging*, 20(S1):72–82. 108, 148

Behnamian, H. and Shafiee, M. (2018). Characterization of three layer depth of interaction PET detectors for small animal imaging. *Biomedical Physics & Engineering Express*, 4(5):055024. 107

Ben Bouallegue, F., Crouzet, J.-F., Comtat, C., Fourcade, M., Mohammadi, B., and Mariano-Goulart, D. (2007). Exact and Approximate Fourier Rebinning Algorithms for the Solution of the Data Truncation Problem in 3-D PET. *IEEE Transactions on Medical Imaging*, 26(7):1001–1009. 18, 20

Berker, Y., Maier, J., and Kachelries, M. (2018). Deep Scatter Estimation in PET: Fast Scatter Correction Using a Convolutional Neural Network. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pages 1–5. IEEE. 33, 36, 137

Berko, S. and Hereford, F. L. (1956). Experimental Studies of Positron Interactions in Solids and Liquids. *Reviews of Modern Physics*, 28(3):299–307. 4

Bert, J., Perez-Ponce, H., Bitar, Z. E., Jan, S., Boursier, Y., Vintache, D., Bonissent, A., Morel, C., Brasse, D., and Visvikis, D. (2013). Geant4-based Monte Carlo simulations on GPU for medical applications. *Physics in Medicine and Biology*, 58(16):5593–5611. 44, 109

Bert, J., Perez-Ponce, H., Jan, S., El Bitar, Z., Gueth, P., Cuplov, V., Chekatt, H., Benoit, D., Sarrut, D., Boursier, Y., Brasse, D., Buvat, I., Morel, C., and Visvikis, D. (2012). Hybrid GATE: A GPU/CPU implementation for imaging and therapy applications. *IEEE Nuclear Science Symposium Conference Record*, pages 2247–2250. 109

Berthon, B., Häggström, I., Apte, A., Beattie, B. J., Kirov, A. S., Humm, J. L., Marshall, C., Spezi, E., Larsson, A., and Schmidtlein, C. R. (2015). PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods. *Physica Medica*, 31(8):969–980. 109

Bettinardi, V., Pagani, E., Gilardi, M., Alenius, S., Thielemans, K., Teras, M., and Fazio, F. (2002). Implementation and evaluation of a 3D one-step late reconstruction algorithm for 3D positron emission tomography brain studies using median root prior. *European Journal of Nuclear Medicine and Molecular Imaging*, 29(1):7–18. 23, 28, 68, 84, 104

Blanc-Durand, P., Khalife, M., Sgard, B., Kaushik, S., Soret, M., Tiss, A., El Fakhri, G., Habert, M.-O., Wiesinger, F., and Kas, A. (2019). Attenuation correction using 3D deep convolutional neural network for brain 18F-FDG PET/MR: Comparison with Atlas, ZTE and CT based attenuation correction. *PLOS ONE*, 14(10):e0223141. 36

Böhlen, T., Cerutti, F., Chin, M., Fassò, A., Ferrari, A., Ortega, P., Mairani, A., Sala, P., Smirnov, G., and Vlachoudis, V. (2014). The FLUKA Code: Developments and Challenges for High Energy and Medical Applications. *Nuclear Data Sheets*, 120:211–214. 108

Bohm, G. and Zech, G. (2014). Statistics of weighted Poisson events and its applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 748:1–6. 103

Bradley, W. G. (2008). History of Medical Imaging. *Proceedings of the American Philosophical Society*, 152(3):349–361. 1

Briesmeister, J. F. (2000). MCNPTM – A General Monte Carlo N-Particle Transport Code. *Los Alamos National Laboratory*. 108

Browne, J. and de Pierro, A. (1996). A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. *IEEE Transactions on Medical Imaging*, 15(5):687–699. xviii, 22, 27

Burger, C., Goerres, G., Schoenes, S., Buck, A., Lonn, A., and von Schulthess, G. (2002). PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. *European Journal of Nuclear Medicine and Molecular Imaging*, 29(7):922–927. 32, 67, 124

Buvat, I. and Lazaro, D. (2006). Monte Carlo simulations in emission tomography and GATE: An overview. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 569(2):323–329. 109

Byars, L., Sibomana, M., Burbar, Z., Jones, J., Panin, V., Barker, W., Liow, J.-S., Carson, R., and Michel, C. (2005). Variance Reduction on Randoms from Delayed Coincidence Histograms for the HRRT. In *IEEE Nuclear Science Symposium Conference Record, 2005*, volume 5, pages 2622–2626. IEEE. 34

Cabello, J., Oliver, J. F., Torres-Espallardo, I., and Rafecas, M. (2010). Polar voxelization schemes combined with a Monte-Carlo based system matrix for image reconstruction in high resolution PET. In *IEEE Nuclear Science Symposuim & Medical Imaging Conference*, pages 3256–3261. IEEE. 61

Cal-Gonzalez (2014). *Positron Range and Promt Gamma Modeling in PET Imaging.* PhD thesis, Universidad Complutense de Madrid. xviii, 118, 119, 143

Cal-González, J., Herraiz, J. L., España, S., Corzo, P. M. G., Vaquero, J. J., Desco, M., and Udias, J. M. (2013). Positron range estimations with PeneloPET. *Physics in Medicine and Biology*, 58(15):5127–5152. xviii, 4, 49, 108, 118, 119, 143

Cal-Gonzalez, J., Herraiz, J. L., Espana, S., Desco, M., Vaquero, J. J., and Udias, J. M. (2009). Positron range effects in high resolution 3D PET imaging. In *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 2788–2791. IEEE. 4, 25, 118

Cal-González, J., Lage, E., Herranz, E., Vicente, E., Udias, J. M., Moore, S. C., Park, M.-A., Dave, S. R., Parot, V., and Herraiz, J. L. (2015a). Simulation of triple coincidences in PET. *Physics in Medicine and Biology*, 60(1):117–136. xviii, 16, 49

Cal-González, J., Pérez-Liva, M., Herraiz, J. L., Vaquero, J. J., Desco, M., and Udías, J. M. (2015b). Tissue-Dependent and Spatially-Variant Positron Range Correction in 3D PET. *IEEE Transactions on Medical Imaging*, 34(11):2394–2403. xviii, 25, 107, 116, 118, 119, 120, 143, 180

Cal-Gonzalez, J., Rausch, I., Shiyam Sundar, L. K., Lassen, M. L., Muzik, O., Moser, E., Papp, L., and Beyer, T. (2018a). Hybrid Imaging: Instrumentation and Data Processing. *Frontiers in Physics*, 6(MAY). xvii, 1, 32

Cal-Gonzalez, J., Vaquero, J. J., Herraiz, J. L., Pérez-Liva, M., Soto-Montenegro, M. L., Peña-Zalbidea, S., Desco, M., and Udías, J. M. (2018b). Improving PET Quantification of Small Animal [^{68}Ga]DOTA-Labeled PET/CT Studies by Using a CT-Based Positron Range Correction. *Molecular Imaging and Biology*, 20(4):584–593. 107, 118

Carter, L. L., Cashwell, E. D., and Taylor, W. M. (1972). Monte Carlo Sampling with Continuously Varying Cross Sections Along Flight Paths. *Nuclear Science and Engineering*, 48(4):403–411. 123

Castiglioni, I., Cremonesi, O., Gilardi, M., Bettinardi, V., Rizzo, G., Savi, A., Bellotti, E., and Fazio, F. (1999). Scatter correction techniques in 3D PET: a Monte Carlo evaluation. *IEEE Transactions on Nuclear Science*, 46(6):2053–2058. 32, 33, 107, 136

Catana, C. (2019a). Development of 7-T MR-compatible TOF-DOI PET Detector and System Technology for the Human Dynamic Neurochemical Connectome Scanner. *[NIH GRANT] Boston, United States: Massachusetts General Hospital*, (1R01EB026995-01). xix, xxii, xxvi, 53, 113, 165

Catana, C. (2019b). Development of Dedicated Brain PET Imaging Devices: Recent Advances and Future Perspectives. *Journal of Nuclear Medicine*, 60(8):1044–1052. xvii, xxii, xxvi, 53, 61, 113, 144

Chaudhari, A. J. and Badawi, R. D. (2021). Application-specific nuclear medical in vivo imaging devices. *Physics in Medicine and Biology*, 66(10):10TR01. 144

Chen, C.-L., Wang, Y., Lee, J. J. S., and Tsui, B. M. W. (2008). Integration of SimSET photon history generator in GATE for efficient Monte Carlo simulations of pinhole SPECT. *Medical Physics*, 35(7Part1):3278–3284. 108

Cheng, X., Hu, K., Yang, D., and Shao, Y. (2021). A compact and lightweight small animal PET with uniform high-resolution for onboard PET/CT image-guided preclinical radiation oncology research. *Physics in Medicine & Biology*, (111):0–13. xvii

Cherry, S., Sorenson, J., and Phelps, M. (2012). *Physics in Nuclear Medicine*. Elsevier. xvii, xxi, xxv, 1, 9, 10, 18, 19, 20, 21, 30, 31, 33, 87, 102, 147, 179

Chi, Y., Tian, Z., and Jia, X. (2016). Modeling parameterized geometry in GPU-based Monte Carlo particle transport simulation for radiotherapy. *Physics in Medicine and Biology*, 61(15):5851–5867. 44, 109

Chien-Min Kao, Yun Dong, and Qingguo Xie (2007). Evaluation of 3D image reconstruction methods for a dual-head small-animal PET scanner. In *2007 IEEE Nuclear Science Symposium Conference Record*, volume 4, pages 3046–3050. IEEE. 25, 148, 160, 161

Choi, H. J., Jang, J. W., Shin, W. G., Park, H., Incerti, S., and Min, C. H. (2020). Development of integrated prompt gamma imaging and positron emission tomography system for in vivo 3-D dose verification: A Monte Carlo study. *Physics in Medicine and Biology*, 65(10). 108

Comtat, C., Bataille, F., Michel, C., Jones, J., Sibomana, M., Janeiro, L., and Trebossen, R. (2004). OSEM-3D Reconstruction Strategies for the ECAT HRRT. In *IEEE Symposium Conference Record Nuclear Science 2004.*, volume 6, pages 3492–3496. IEEE. 24, 147

Comtat, C., Kinahan, P., Defrise, M., Michel, C., Lartizien, C., and Townsend, D. (2003). Simulating whole-body PET scanning with rapid analytical methods. In *1999 IEEE Nuclear Science Symposium. Conference Record. 1999 Nuclear Science Symposium and Medical Imaging Conference (Cat. No.99CH37019)*, volume 3, pages 1260–1264. IEEE. 109

Conti, M. (2011). Focus on time-of-flight PET: the benefits of improved time resolution. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(6):1147–1157. 14, 90, 106

Corda-D'Incan, G., Schnabel, J. A., and Reader, A. J. (2020a). Iteration-Dependent Networks and Losses for Unrolled Deep Learned FBSEM PET Image Reconstruction. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4. IEEE. 36

Corda-D'Incan, G., Schnabel, J. A., and Reader, A. J. (2020b). Syn-Net for Synergistic Deep-Learned PET-MR Reconstruction. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–5. IEEE. 36

Cui, J., Gong, K., Guo, N., Wu, C., Kim, K., Liu, H., and Li, Q. (2021). Populational and individual information based PET image denoising using conditional unsupervised learning. *Physics in Medicine & Biology*, 66(15):155001. 36

Cui, J., Gong, K., Guo, N., Wu, C., Meng, X., Kim, K., Zheng, K., Wu, Z., Fu, L., Xu, B., Zhu, Z., Tian, J., Liu, H., and Li, Q. (2019). PET image denoising using unsupervised deep learning. *European Journal of Nuclear Medicine and Molecular Imaging*, 46(13):2780–2789. 36

Daube-Witherspoon, M. E. and Muehllehner, G. (1986). An Iterative Image Space Reconstruction Algorthm Suitable for Volume ECT. *IEEE Transactions on Medical Imaging*, 5(2):61–66. 22

Daube-Witherspoon, M. E. and Muehllehner, G. (1987). Treatment of axial data in three-dimensional PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 28(11):1717–24. 17, 20

Davis, K. M., Ryan, J. L., Aaron, V. D., and Sims, J. B. (2020). PET and SPECT Imaging of the Brain: History, Technical Considerations, Applications, and Radiotracers. *Seminars in Ultrasound, CT and MRI*, 41(6):521–529. 1

De Pierro, A. (1995). A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging*, 14(1):132–137. 23, 27, 28

De Pierro, A. and Yamagishi, M. (2001). Fast EM-like methods for maximum "a posteriori" estimates in emission tomography. *IEEE Transactions on Medical Imaging*, 20(4):280–288. 23, 27, 28

DeBenedetti, S., Cowan, C. E., Konneker, W. R., and Primakoff, H. (1950). On the Angular Distribution of Two-Photon Annihilation Radiation. *Physical Review*, 77(2):205–212. 5

Defrise, M., Kinahan, P., Townsend, D., Michel, C., Sibomana, M., and Newport, D. (1997). Exact and approximate rebinning algorithms for 3-D PET data. *IEEE Transactions on Medical Imaging*, 16(2):145–158. 17, 20

Defrise, M., Townsend, D. W., Bailey, D., Geissbuhler, A. M. C., and Jones, T. (1991). A normalization technique for 3D PET data. *Physics in Medicine and Biology*, 36(7):939–952. 30, 66

Deidda, D., Karakatsanis, N. A., Robson, P. M., Tsai, Y.-J., Efthimiou, N., Thielemans, K., Fayad, Z. A., Aykroyd, R. G., and Tsoumpas, C. (2019). Hybrid PET-MR list-mode kernelized expectation maximization reconstruction. *Inverse Problems*, 35(4):044001. 27

Delso, G., Furst, S., Jakoby, B., Ladebeck, R., Ganter, C., Nekolla, S. G., Schwaiger, M., and Ziegler, S. I. (2011). Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner. *Journal of Nuclear Medicine*, 52(12):1914–1922. xxii, xxvi, 10, 17, 27, 32, 52, 87, 112

Delso, G., Martinez, M.-J., Torres, I., Ladebeck, R., Michel, C., Nekolla, S., and Ziegler, S. I. (2009). Monte Carlo simulations of the count rate performance of a clinical whole-body MR/PET scanner. *Medical Physics*, 36(9Part1):4126–4135. 107

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22. 22, 25, 67

Derenzo, S. E., Budinger, T. F., Cahoon, J. L., Huesman, R. H., and Jackson, H. G. (1977). High Resolution Computed Tomography of Positron Emitters. *IEEE Transactions on Nuclear Science*, 24(1):544–558. 58

Després, P. and Jia, X. (2017). A review of GPU-based medical image reconstruction. *Physica Medica*, 42(October 2017):76–92. 44

do Nascimento, E., Vanin, V. R., Maidana, N. L., and Helene, O. (2013). Coincidence Doppler Broadening of Positron Annihilation Radiation in Fe. *Journal of Physics: Conference Series*, 443(1):012024. 5

Dong, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Curran, W. J., Mao, H., Nye, J. A., and Yang, X. (2019). Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. *Physics in Medicine & Biology*, 64(21):215016. 36

Du, Y., Frey, E. C., Wang, W. T., Tocharoenchai, C., Baird, W. H., and Tsui, B. M. (2002). Combination of MCNP and SimSET for Monte Carlo simulation of SPECT with medium- and high-energy photons. *IEEE Transactions on Nuclear Science*, 49 I(3):668–674. 108

Easton, R. L. (2010). *Fourier Methods in Imaging.* John Wiley & Sons, Ltd, Chichester, UK. 19

Eklund, A., Dufort, P., Forsberg, D., and LaConte, S. M. (2013). Medical image processing on the GPU – Past, present and future. *Medical Image Analysis*, 17(8):1073–1094. 44

Erdman, K. L. (1955). The Angular Correlation of Annihilation Radiation. *Proceedings of the Physical Society. Section A*, 68(4):304–311. 5

España, S. (2009). *Simulaciones avanzadas aplicadas al diseño de escáneres y mejora de la calidad de imagen en tomografía por emisión de positrones.* PhD thesis, Universidad Complutense de Madrid. xviii

España, S., Herraiz, J. L., Vicente, E., Herranz, E., Vaquero, J. J., Deseo, M., and Udias, J. M. (2007). Improved image reconstruction in small animal PET using a priori estimates of single-pixel events. *IEEE Nuclear Science Symposium Conference Record*, 5:3876–3880. xviii, 25

España, S., Herraiz, J. L., Vicente, E., Vaquero, J. J., Desco, M., and Udias, J. M. (2009). PeneloPET, a Monte Carlo PET simulation tool based on PENELOPE: features and validation. *Physics in Medicine and Biology*, 54(6):1723–1742. xviii, xxii, xxvii, 49, 108

Fatica, M. and Ruetsch, G. (2014). *CUDA Fortran for Scientists and Engineers.* Elsevier. 44, 45, 117, 142

Feng, T., Yao, S., Xi, C., Zhao, Y., Wang, R., Wu, S., Li, C., and Xu, B. (2021). Deep learning-based image reconstruction for TOF PET with DIRECT data partitioning format. *Physics in Medicine & Biology*, 66(16):165007. 36

Ferrari, A., Sala, P. R., Fassò, A., and Ranft, J. (2005). FLUKA: A Multi-Particle Transport Code. Technical report, CERN. 108

Freese, D. L., Olcott, P. D., Buss, S. R., and Levin, C. S. (2018). Gray: a ray tracing-based Monte Carlo simulator for PET. *Physics in Medicine and Biology*, 63(10):105019. 108

Freire, M., Gonzalez-Montoro, A., Canizares, G., Rezaei, A., Nuyts, J., Berr, S. S., Williams, M. B., Benlloch, J. M., and Gonzalez, A. J. (2021). Experimental validation of a rodent PET scanner prototype based on a single LYSO crystal tube. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 7311(c):1–1. xvii, 12

Galve, P., Arias-Valcayo, F., Lopez-Montes, A., Villa-Abaunza, A., Ibanez, P., Herraiz, J. L., and Udias, J. M. (2020a). Multi-purpose Ultra-fast Monte Carlo PET simulator. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC).* 44, 110

Galve, P., Arias-Valcayo, F., Lopez-Montes, A., Villa-Abaunza, A., Ibanez, P., Herraiz, J. L., and Udias, J. M. (2021a). Ultra-fast Fully Monte Carlo PET Reconstruction. In *2021 Virtual IEEE Nuclear Science Symposium and Medical Imaging Conference.* 148

Galve, P., Arias-Valcayo, F., Lopez-Montes, A., Villa-Abaunza, A., Ibanez, P., Herraiz, J. L., and Udias, J. M. (2021b). Ultra-fast Monte Carlo PET Reconstructor. In *16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, pages 152–156. 148

Galve, P., Herraiz, J. L., Catana, C., and Udias, J. M. (2020b). GPU based Fast and Flexible Iterative Reconstructions of Arbitrary and Complex PET Scanners: Application to Next Generation Dedicated Brain Scanners. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. 23, 53, 62, 113, 144

Galve, P., Herraiz, J. L., Lopez-Montes, A., Gutierrez, C., and Udias, J. M. (2018). Effect of Incomplete Acquisitions on the Image of High Resolution Preclinical Scanners. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, Sidney, Australia. 82

Galve, P., López-Montes, A., Udías, J. M., and López Herraiz, J. (2019a). Iterative Algorithm for Optimal Super Resolution Sampling. In *Springer Proceedings in Physics*, volume 225, pages 141–143. 89

Galve, P., Lopez-Montes, A., Udias, J. M., Moore, S. C., and Herraiz, J. L. (2017). Data-driven Improved Sampling in PET. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–5. IEEE. 89, 103

Galve, P., Udias, J. M., Lopez-Montes, A., Arias-Valcayo, F., Vaquero, J. J., Desco, M., and Herraiz, J. L. (2021c). Super-Iterative Image Reconstruction in PET. *IEEE Transactions on Computational Imaging*, 7(c):248–257. xxii, xxvi, 82, 89

Galve, P., Udias, J. M., Lopez-Montes, A., and Herraiz, J. L. (2019b). Super-iterative image reconstruction in PET. In Matej, S. and Metzler, S. D., editors, *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, number May, page 53. SPIE. 89

Garcia, M.-P., Bert, J., Benoit, D., Bardiès, M., and Visvikis, D. (2016). Accelerated GPU based SPECT Monte Carlo simulations. *Physics in Medicine and Biology*, 61(11):4001–4018. 109

Gillam, J. E. and Rafecas, M. (2016). Monte-Carlo simulations and image reconstruction for novel imaging scenarios in emission tomography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 809:76–88. 24, 107, 148

Gong, K., Kim, K., Cui, J., Wu, D., and Li, Q. (2021). The Evolution of Image Reconstruction in PET: From Filtered Back-Projection to Artificial Intelligence. *PET Clinics*, 16(4):533–542. xviii, 18, 36

Gong, K., Wu, D., Kim, K., Yang, J., Sun, T., El Fakhri, G., Seo, Y., and Li, Q. (2019). MAPEM-Net: an unrolled neural network for Fully 3D PET image reconstruction. In Matej, S. and Metzler, S. D., editors, *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, page 102. SPIE. 36

Gong, K., Zhou, J., Tohme, M., Judenhofer, M., Yang, Y., and Qi, J. (2017). Sinogram Blurring Matrix Estimation From Point Sources Measurements With Rank-One Approximation for Fully 3-D PET. *IEEE Transactions on Medical Imaging*, 36(10):2179–2188. 25, 61, 88, 147, 161

Gordon, R., Bender, R., and Herman, G. T. (1970). Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481. 22

Green, P. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93. xviii, 23, 27, 28, 29, 84

Grkovski, M., Brzezinski, K., Cindro, V., Clinthorne, N. H., Kagan, H., Lacasta, C., Mikuž, M., Solaz, C., Studen, A., Weilhammer, P., and Žontar, D. (2015). Evaluation of a high resolution silicon PET insert module. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 788:86–94. xvii, 10, 144

Guerin, B. and El Fakhri, G. (2008). Realistic PET Monte Carlo Simulation With Pixelated Block Detectors, Light Sharing, Random Coincidences and Dead-Time Modeling. *IEEE Transactions on Nuclear Science*, 55(3):942–952. 108

Guo, H. and Renaut, R. A. (2011). Revisiting stopping rules for iterative methods used in emission tomography. *Computerized Medical Imaging and Graphics*, 35(5):398–406. 28

Guobao Wang and Jinyi Qi (2012). Penalized Likelihood PET Image Reconstruction Using Patch-Based Edge-Preserving Regularization. *IEEE Transactions on Medical Imaging*, 31(12):2194–2204. 28

Häggström, I., Schmidtlein, C. R., Campanella, G., and Fuchs, T. J. (2019). DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical Image Analysis*, 54:253–262. 36

Hallen, P., Schug, D., and Schulz, V. (2020). Comments on the NEMA NU 4-2008 Standard on Performance Measurement of Small Animal Positron Emission Tomographs. *EJNMMI Physics*, 7(1):12. 58

Herraiz, J., España, S., Vicente, E., Vaquero, J., Desco, M., and Udias, J. (2007). Noise and physical limits to maximum resolution of PET images. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 580(2):934–937. 90

Herraiz, J. L., Bembibre, A., and López-Montes, A. (2020). Deep-Learning Based Positron Range Correction of PET Images. *Applied Sciences*, 11(1):266. 36, 143

Herraiz, J. L., España, S., Cabido, R., Montemayor, A. S., Desco, M., Vaquero, J. J., and Udias, J. M. (2011). GPU-Based Fast Iterative Reconstruction of Fully 3-D PET Sinograms. *IEEE Transactions on Nuclear Science*, 58(5):2257–2263. xviii, 23, 61, 83

Herraiz, J. L., España, S., Vaquero, J. J., Desco, M., and Udías, J. M. (2006a). FIRST: Fast Iterative Reconstruction Software for (PET) tomography. *Physics in Medicine and Biology*, 51(18):4547–4565. xviii, 25, 27, 61, 88, 90, 105, 106, 107, 148, 161

Herraiz, J. L., Espana, S., Vicente, E., Herranz, E., Desco, M., Vaquero, J. J., and Udias, J. (2008). Frequency selective signal extrapolation for compensation of missing data in sinograms. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 4299–4302. IEEE. xviii, 20, 88

Herraiz, J. L., Espana, S., Vicente, E., Vaquero, J. J., Desco, M., and Udias, J. M. (2006b). Optimal and Robust PET Data Sinogram Restoration Based on the Response of the System. In *2006 IEEE Nuclear Science Symposium Conference Record*, volume 6, pages 3404–3407. IEEE. 20, 88

Heußer, T., Mann, P., Rank, C. M., Schäfer, M., Dimitrakopoulou-Strauss, A., Schlemmer, H.-P., Hadaschik, B. A., Kopka, K., Bachert, P., Kachelrieß, M., and Freitag, M. T. (2017). Investigation of the halo-artifact in 68Ga-PSMA-11-PET/MRI. *PLOS ONE*, 12(8):e0183329. xix, 32, 33, 136

Hissoiny, S., Ozell, B., Bouchard, H., and Després, P. (2011). GPUMCD: A new GPU-oriented Monte Carlo dose calculation platform. *Medical Physics*, 38(2):754–764. 44, 109, 123, 142, 143

Hoffman, E., Guerrero, T., Germano, G., Digby, W., and Dahlbom, M. (1989). PET system calibrations and corrections for quantitative and spatially accurate images. *IEEE Transactions on Nuclear Science*, 36(1):1108–1112. 30, 66

Holdsworth, C., Levin, C., Farquhar, T., Dahlbom, M., and Hoffman, E. (2001). Investigation of accelerated Monte Carlo techniques for PET simulation and 3D PET scatter correction. *IEEE Transactions on Nuclear Science*, 48(1):74–81. 33, 136

Hong, X., Zan, Y., Weng, F., Tao, W., Peng, Q., and Huang, Q. (2018). Enhancing the Image Quality via Transferred Deep Residual Learning of Coarse PET Sinograms. *IEEE Transactions on Medical Imaging*, 37(10):2322–2332. 89, 103

Hudson, H. and Larkin, R. (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609. xviii, xix, xxi, xxiii, xxv, xxvii, 23, 27, 62, 67, 104, 150, 159

Huesman, R., Klein, G., Moses, W., Jinyi Qi, Reutter, B., and Virador, P. (2000). List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling. *IEEE Transactions on Medical Imaging*, 19(5):532–537. 24, 88, 104

Hwang, D., Kang, S. K., Kim, K. Y., Seo, S., Paeng, J. C., Lee, D. S., and Lee, J. S. (2019). Generation of PET Attenuation Map for Whole-Body Time-of-Flight 18 F-FDG PET/MRI Using a Deep Neural Network Trained with Simultaneously Reconstructed Activity and Attenuation Maps. *Journal of Nuclear Medicine*, 60(8):1183–1189. 36

Ibáñez, P., Villa-Abaunza, A., Vidal, M., Guerra, P., Graullera, S., Illana, C., and Udías, J. M. (2021). XIORT-MC: A real-time MC-based dose computation tool for low- energy X-rays intraoperative radiation therapy. *Medical Physics*, 48(12):8089–8106. 44

Iriarte, A., Marabini, R., Matej, S., Sorzano, C., and Lewitt, R. (2016). System models for PET statistical iterative reconstruction: A review. *Computerized Medical Imaging and Graphics*, 48:30–48. xviii, 18, 22, 23, 24, 27, 63, 88, 147

Izquierdo-Garcia, D. and Catana, C. (2016). MR Imaging–Guided Attenuation Correction of PET Data in PET/MR Imaging. *PET Clinics*, 11(2):129–149. 32

Jahnke, L., Fleckenstein, J., Wenz, F., and Hesser, J. (2012). GMC: a GPU implementation of a Monte Carlo dose calculation based on Geant4. *Physics in Medicine and Biology*, 57(5):1217–1229. 109

Jakoby, B., Long, M., Carr, C., and Townsend, D. (2007). Physical performance of a new combined PET/CT scanner. *Journal of Nuclear Medicine*, 48(supplement 2):46P–46P. 87

Jakoby, B. W., Bercier, Y., Watson, C. C., Bendriem, B., and Townsend, D. W. (2009). Performance Characteristics of a New LSO PET/CT Scanner With Extended Axial Field-of-View and PSF Reconstruction. *IEEE Transactions on Nuclear Science*, 56(3):633–639. 10, 87

James, F. (1990). A review of pseudorandom number generators. *Computer Physics Communications*, 60(3):329–344. 38, 118

Jan, S., Benoit, D., Becheva, E., Carlier, T., Cassol, F., Descourt, P., Frisson, T., Grevillot, L., Guigues, L., Maigne, L., Morel, C., Perrot, Y., Rehfeld, N., Sarrut, D., Schaart, D. R., Stute, S., Pietrzyk, U., Visvikis, D., Zahra, N., and Buvat, I. (2011). GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Physics in Medicine and Biology*, 56(4):881–901. 108

Jan, S., Frisson, T., and Sarrut, D. (2013). GATE Simulation of ^{12}C Hadrontherapy Treatment Combined With a PET Imaging System for Dose Monitoring: A Feasibility Study. *IEEE Transactions on Nuclear Science*, 60(1):423–429. 108

Jan, S., Santin, G., Strul, D., Staelens, S., Assié, K., Autret, D., Avner, S., Barbier, R., Bardiès, M., Bloomfield, P. M., Brasse, D., Breton, V., Bruyndonckx, P., Buvat, I., Chatziioannou, A. F., Choi, Y., Chung, Y. H., Comtat, C., Donnarieix, D., Ferrer, L., Glick, S. J., Groiselle, C. J., Guez, D., Honore, P.-F., Kerhoas-Cavata, S., Kirov, A. S., Kohli, V., Koole, M., Krieguer, M., van der Laan, D. J., Lamare, F., Largeron, G., Lartizien, C., Lazaro, D., Maas,

M. C., Maigne, L., Mayet, F., Melot, F., Merheb, C., Pennacchio, E., Perez, J., Pietrzyk, U., Rannou, F. R., Rey, M., Schaart, D. R., Schmidtlein, C. R., Simon, L., Song, T. Y., Vieira, J.-M., Visvikis, D., de Walle, R. V., Wieërs, E., and Morel, C. (2004). GATE: a simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*, 49(19):4543–4561. 108

Jeong, K. Y., Choi, K., Nam, W. H., and Ra, J. B. (2011). Sinogram-based super-resolution in PET. *Physics in Medicine and Biology*, 56(15):4881–4894. 88, 104

Jeong, K. Y., Kim, J. H., Kim, K. M., and Ra, J. B. (2013). Sinogram Super-Resolution Using a Space-Variant Blur Matrix in PET. *IEEE Transactions on Nuclear Science*, 60(1):158–165. 88, 104

Jia, X., Gu, X., Graves, Y. J., Folkerts, M., and Jiang, S. B. (2011). GPU-based fast Monte Carlo simulation for radiotherapy dose calculation. *Physics in Medicine and Biology*, 56(22):7017–7031. 44, 109, 142

Jia, X., Gu, X., Sempau, J., Choi, D., Majumdar, A., and Jiang, S. B. (2010). Development of a GPU-based Monte Carlo dose calculation code for coupled electron–photon transport. *Physics in Medicine and Biology*, 55(11):3077–3086. 109

Jia, X., Schümann, J., Paganetti, H., and Jiang, S. B. (2012a). GPU-based fast Monte Carlo dose calculation for proton therapy. *Physics in Medicine and Biology*, 57(23):7783–7797. 109

Jia, X., Yan, H., Gu, X., and Jiang, S. B. (2012b). Fast Monte Carlo simulation for patient-specific CT/CBCT imaging dose calculation. *Physics in Medicine and Biology*, 57(3):577–590. 109

Jin, X., Chan, C., Mulnix, T., Liu, C., and Carson, R. E. (2012). List-mode reconstruction for the Biograph mCT with probabilistic line-of-response positioning and event-by-event motion correction. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, pages 3682–3687. IEEE. 89, 103, 104

Kadrmas, D. J. (2004). LOR-OSEM: statistical PET reconstruction from raw line-of-response histograms. *Physics in Medicine and Biology*, 49(20):4731–4744. 16, 69

Kahn, H. (1954). APPLICATIONS OF MONTE CARLO. Technical report, RAND Corp., Santa Monica, Calif., United States. 40

Kalos, M. H. and Whitlock, P. A. (2008). *Monte Carlo Methods.* Wiley, Germany, Darmstadt, second, re edition. 37, 38, 39, 40, 43

Kang, H. G., Tashima, H., Nishikido, F., Akamatsu, G., Wakizaka, H., Higuchi, M., and Yamaya, T. (2021). Initial results of a mouse brain PET insert with a staggered 3-layer DOI detector. *Physics in Medicine & Biology*, 2(1):0–31. xvii, 10, 24, 32, 61, 65

Karp, J., Muehllehner, G., and Lewitt, R. (1988). Constrained Fourier space method for compensation of missing data in emission computed tomography. *IEEE Transactions on Medical Imaging*, 7(1):21–25. 88

Kawrakow, I. (2000a). Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version. *Medical Physics*, 27(3):485–498. 108

Kawrakow, I. (2000b). Accurate condensed history Monte Carlo simulation of electron transport. II. Application to ion chamber response simulations. *Medical Physics*, 27(3):499–513. 108

Kennedy, J., Israel, O., Frenkel, A., Bar-Shalom, R., and Haim Azhari (2006). Super-resolution in PET imaging. *IEEE Transactions on Medical Imaging*, 25(2):137–147. 88, 104

Kim, K., Dutta, J., Groll, A., El Fakhri, G., Meng, L.-J., and Li, Q. (2018). A novel depth-of-interaction rebinning strategy for ultrahigh resolution PET. *Physics in Medicine & Biology*, 63(16):165011. 89, 103

Kinahan, P. and Rogers, J. (1989). Analytic 3D image reconstruction using all detected events. *IEEE Transactions on Nuclear Science*, 36(1):964–968. 20

Knoll, G. F. (2011). *Radiation Detection and Measurements.* John Willey & Son, fourth edi edition. 4, 5, 7, 8, 9, 10, 11, 13, 14, 34, 35, 169

Kotasidis, F. A., Matthews, J. C., Angelis, G. I., Noonan, P. J., Jackson, A., Price, P., Lionheart, W. R., and Reader, A. J. (2011). Single scan parameterization of space-variant point spread functions in image space via a printed array: the impact for two PET/CT scanners. *Physics in Medicine and Biology*, 56(10):2917–2942. 25, 59, 64, 83, 88, 147

Kraan, A. C. (2015). Range Verification Methods in Particle Therapy: Underlying Physics and Monte Carlo Modeling. *Frontiers in Oncology*, 5(JUN):1–27. 107

Krane, K. S. (1988). *Introductory nuclear physics.* Wiley, New York, NY, US. 3, 4, 169

Kraus, R., Delso, G., and Ziegler, S. I. (2012). Simulation study of tissue-specific positron range correction for the new biograph mMR whole-body PET/MR system. *IEEE Transactions on Nuclear Science*, 59(5 PART 1):1900–1909. 25, 107

Kyung Sang Kim, Young Don Son, Zang Hee Cho, Jong Beom Ra, and Jong Chul Ye (2014). Ultra-Fast Hybrid CPU–GPU Multiple Scatter Simulation for 3-D PET. *IEEE Journal of Biomedical and Health Informatics*, 18(1):148–156. 33, 136

Ladefoged, C. N., Law, I., Anazodo, U., St. Lawrence, K., Izquierdo-Garcia, D., Catana, C., Burgos, N., Cardoso, M. J., Ourselin, S., Hutton, B., Mérida, I., Costes, N., Hammers, A., Benoit, D., Holm, S., Juttukonda, M., An, H., Cabello, J., Lukas, M., Nekolla, S., Ziegler, S., Fenchel, M., Jakoby, B., Casey, M. E., Benzinger, T., Højgaard, L., Hansen, A. E., and Andersen, F. L. (2017). A multi-centre evaluation of eleven clinically feasible brain PET/MRI attenuation correction techniques using a large cohort of patients. *NeuroImage*, 147(December 2016):346–359. 32

Lage, E., Parot, V., Moore, S. C., Sitek, A., Udías, J. M., Dave, S. R., Park, M.-A., Vaquero, J. J., and Herraiz, J. L. (2015). Recovery and normalization of triple coincidences in PET. *Medical Physics*, 42(3):1398–1410. xviii, 16, 89, 91, 103

Lai, Y., Zhong, Y., Chalise, A., Shao, Y., Jin, M., Jia, X., and Chi, Y. (2019). gPET: a GPU-based, accurate and efficient Monte Carlo simulation tool for PET. *Physics in Medicine and Biology*, 64(24):245002. 44, 109, 142, 143, 145

Laurrent, B., Merlin, T., Bousse, A., and Visvikis, D. (2020). Deep learning based scatter correction for PET imaging. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*. 33, 36, 137

Lecoq, P., Morel, C., Prior, J. O., Visvikis, D., Gundacker, S., Auffray, E., Križan, P., Turtos, R. M., Thers, D., Charbon, E., Varela, J., de La Taille, C., Rivetti, A., Breton, D., Pratte, J.-F., Nuyts, J., Surti, S., Vandenberghe, S., Marsden, P., Parodi, K., Benlloch, J. M., and Benoit, M. (2020). Roadmap toward the 10 ps time-of-flight PET challenge. *Physics in Medicine & Biology*, 65(21):21RM01. xvii, 11

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. *Communications of the ACM*, 31(6):742–751. 38, 118

Lee, J. S. (2021). A Review of Deep-Learning-Based Approaches for Attenuation Correction in Positron Emission Tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(2):160–184. 36

Lee, M. S., Kang, S. K., and Lee, J. S. (2018). Novel inter-crystal scattering event identification method for PET detectors. *Physics in Medicine and Biology*, 63:115015. 89, 103, 107

Lewellen, T. K. (2008). Recent developments in PET detector technology. *Physics in Medicine and Biology*, 53(17):R287–R317. 11, 12, 179

Li, K., Safavi-Naeini, M., Franklin, D. R., Han, Z., Rosenfeld, A. B., Hutton, B., and Lerch, M. L. F. (2015). A new virtual ring-based system matrix generator for iterative image reconstruction in high resolution small volume PET systems. *Physics in Medicine and Biology*, 60(17):6949–6973. 24, 25, 61, 107, 148

Li, K., Safavi-Naeini, M., Franklin, D. R., Petasecca, M., Guatelli, S., Rosenfeld, A. B., Hutton, B. F., and Lerch, M. L. F. (2013). A feasibility study of PETiPIX: an ultra high resolution small animal PET scanner. *Journal of Instrumentation*, 8(12):P12004–P12004. 107

Li, T., Zhang, M., Qi, W., Asma, E., and Qi, J. (2020). Motion correction of respiratory-gated PET images using deep learning based image registration framework. *Physics in Medicine & Biology*, 65(15):155003. 36

Li, Y., Matej, S., and Metzler, S. D. (2014). Image reconstructions from super-sampled data sets with resolution modeling in PET imaging. *Medical physics*, 41(12):121912. 88, 104

Lin, H.-H., Chuang, K.-S., Lin, Y.-H., Ni, Y.-C., Wu, J., and Jan, M.-L. (2014). Efficient simulation of voxelized phantom in GATE with embedded SimSET multiple photon history generator. *Physics in Medicine and Biology*, 59(20):6231–6250. 108

Lindemann, M. E., Guberina, N., Wetter, A., Fendler, W. P., Jakoby, B., and Quick, H. H. (2019). Improving 68 Ga-PSMA PET/MRI of the Prostate with Unrenormalized Absolute Scatter Correction. *Journal of Nuclear Medicine*, 60(11):1642–1648. xix, 32, 33, 136

Lippuner, J. and Elbakri, I. A. (2011). A GPU implementation of EGSnrc's Monte Carlo photon transport for imaging applications. *Physics in Medicine and Biology*, 56(22):7145–7162. 44, 109

Ljungberg, M., Strand, S.-E., and King, M. A. (2012). *Monte Carlo Calculations in Nuclear Medicine*. CRC Press, Boca Raton, 2nd edition edition. 12, 37

Lopez-Montes, A. (2021). *Improved Techniques for PET Imaging.* PhD thesis, Universidad Complutense de Madrid. xviii, 16, 18, 110, 145

Lopez-Montes, A., Galve, P., Udias, J., and Herraiz, J. (2017). Real-Time Accurate Rebinning of PET Data Based on the Pseudo-Inverse of the Axial System Matrix. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, volume 2, pages 1–4. IEEE. 18, 20

López-Montes, A., Galve, P., Udias, J. M., Cal-González, J., Vaquero, J. J., Desco, M., and Herraiz, J. L. (2020). Real-Time 3D PET Image with Pseudoinverse Reconstruction. *Applied Sciences*, 10(8):2829. xviii, 18, 20

Lopez Montes, A., Galve, P., Udias, J. M., and Herraiz, J. L. (2019). Application of the pseudoinverse for real-time 3D PET image reconstruction. In Matej, S. and Metzler, S. D., editors, *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, number May, page 74. SPIE. 18, 20

Lopez-Montes, A., Herraiz, J. L., Galve, P., Espana, S., Vicente, E., Cal-Gonzalez, J., and Udias, J. M. (2019). PeneloPET v3.0, an improved multiplatform PET Simulator. In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3. IEEE. xviii, xxii, xxvii, 49, 50, 108

Lougovski, A., Hofheinz, F., Maus, J., Schramm, G., Will, E., and Hoff, J. V. D. (2014). A volume of intersection approach for on-the-fly system matrix calculation in 3D PET image reconstruction. *Physics in Medicine and Biology*, 59(3):561–577. 24

Ma, B., Xu, H., Lenz, M., Pietrzyk, U., Shah, N. J., Gaens, M., Caldeira, L., Bert, J., Lohmann, P., Tellmann, L., Lerche, C., Scheins, J., and Rota Kops, E. (2020). Scatter Correction Based on GPU-Accelerated Full Monte Carlo Simulation for Brain PET/MRI. *IEEE Transactions on Medical Imaging*, 39(1):140–151. 33, 44, 107, 109, 136, 137, 142, 143, 145

Marcatili, S., Collot, J., Curtoni, S., Dauvergne, D., Hostachy, J.-Y., Koumeir, C., Létang, J. M., Livingstone, J., Métivier, V., Gallin-Martel, L., Gallin-Martel, M. L., Muraz, J. F., Servagent, N., Testa, É., and Yamouni, M. (2020). Ultra-fast prompt gamma detection in single proton counting regime for range monitoring in particle therapy. *Physics in Medicine & Biology*, 65(24):245033. 107

Markiewicz, P. J., Thielemans, K., Ehrhardt, M. J., Jiao, J., Burgos, N., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S. (2014). High throughput CUDA implementation of accurate geometric modelling for iterative reconstruction of PET data. In *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4. IEEE. 24, 61, 90, 161

Marsaglia, G., MacLaren, M. D., and Bray, T. A. (1964). A fast procedure for generating normal random variables. *Communications of the ACM*, 7(1):4–10. 40

Marsaglia, G., Narasimhan, B., and Zaman, A. (1990). A random number generator for PC's. *Computer Physics Communications*, 60(3):345–349. 38

Masuda, T., Nishio, T., Kataoka, J., Arimoto, M., Sano, A., and Karasawa, K. (2019). ML-EM algorithm for dose estimation using PET in proton therapy. *Physics in Medicine & Biology*, 64(17):175011. 107

Masuda, T., Nishio, T., Sano, A., and Karasawa, K. (2020). Extension of the ML-EM algorithm for dose estimation using PET in proton therapy: application to an inhomogeneous target. *Physics in Medicine & Biology*, 64(17):175011. 107

McIntyre, R. (1985). Recent developments in silicon avalanche photodiodes. *Measurement*, 3(4):146–152. 13

Meikle, S. R., Dahlbom, M., and Cherry, S. R. (1993). Attenuation correction using count-limited transmission data in positron emission tomography. *Journal of Nuclear Medicine*, 34(1):143–144. 31

Meng, F., Shi, Y., Li, C., Li, L., Qin, W., and Zhu, S. (2021). Hybrid model of photon propagation based on the analytical and Monte Carlo methods for a dual-head PET system. *Physics in Medicine & Biology*, 66(17):175008. 25, 105, 148

Meng, F., Wang, J., Zhu, S., Cheng, J., Liang, J., and Tian, J. (2019). Comparison of GPU reconstruction based on different symmetries for dual-head PET. *Medical Physics*, 46(6):2696–2708. 23, 25, 61, 83, 148, 160, 161

Michel, C., Schmand, M., Liu, X., Sibomana, M., Vollmar, S., Knoss, C., Lercher, M., Watson, C., Newport, D., Casey, M., Defrise, M., Wienhard, K., and Heiss, W. (2000). Reconstruction strategies for the HRRT. In *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*, volume 2, pages 15/207–15/212. IEEE. 24, 147

Michel Defrise and Paul Kinahan (1998). Data Acquisition and Image Reconstruction for 3D PET. In Bendriem, B. and Townsend, D. W., editors, *The Theory and Practice of 3D PET*, chapter Chapter 2, pages 11–53. Kluwer Academic Publisher. xvii, 16, 17, 21

Miranda, A., Bertoglio, D., Glorie, D., Stroobants, S., Staelens, S., and Verhaeghe, J. (2020). Validation of a spatially variant resolution model for small animal brain PET studies. *Biomedical Physics & Engineering Express*, 6(4):045001. 25, 147

Moehrs, S., Defrise, M., Belcari, N., Guerra, A. D., Bartoli, A., Fabbri, S., and Zanetti, G. (2008). Multi-ray-based system matrix generation for 3D PET reconstruction. *Physics in Medicine and Biology*, 53(23):6925–6945. 24, 88, 104

Mohammadi, A., Inadama, N., Yoshida, E., Nishikido, F., Shimizu, K., and Yamaya, T. (2017). Improvement of crystal identification performance for a four-layer DOI detector composed of crystals segmented by laser processing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 866(October 2016):29–35. xvii, 61, 87, 115, 144

Mohammadi, I., Castro, I. F. C., Correia, P. M. M., Silva, A. L. M., and Veloso, J. F. C. A. (2019). Minimization of parallax error in positron emission tomography using depth of interaction capable detectors: methods and apparatus. *Biomedical Physics & Engineering Express*, 5(6):062001. 24, 147

Mok, G. S. P., Du, Y., Wang, Y., Frey, E. C., and Tsui, B. M. W. (2010). Development and Validation of a Monte Carlo Simulation Tool for Multi-Pinhole SPECT. *Molecular Imaging and Biology*, 12(3):295–304. 108

Moore, G. E. (2006). Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Society Newsletter*, 11(3):33–35. 44

Moore, S., Cervo, M., Udias, J. M., Herraiz, J. L., and Metzler, S. (2015). An Iterative Method for Eliminating Artifacts from Multiplexed Data in Pinhole SPECT. In *Fully3D conference*, pages 515–517. 89, 103

Morales, M. A., Izquierdo-Garcia, D., Aganj, I., Kalpathy-Cramer, J., Rosen, B. R., and Catana, C. (2019). Implementation and Validation of a Three-dimensional Cardiac Motion Estimation Network. *Radiology: Artificial Intelligence*, 1(4):e180080. 36

Moses, W. W. (2011). Fundamental limits of spatial resolution in PET. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648:S236–S240. 87, 90, 91, 102

Mumcuoglu, E. Ü., Leahy, R. M., and Cherry, S. R. (1996). Bayesian reconstruction of PET images: methodology and performance analysis. *Physics in Medicine and Biology*, 41(9):1777–1807. 28

Munk, O. L., Tolbod, L. P., Hansen, S. B., and Bogsrud, T. V. (2017). Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative. *EJNMMI Physics*, 4(1):5. 84

Nadig, V., Herrmann, K., Mottaghy, F. M., and Schulz, V. (2021). Hybrid total-body pet scanners—current status and future perspectives. *European Journal of Nuclear Medicine and Molecular Imaging*. xvii, 10, 32, 61

Nagy, K., Tóth, M., Major, P., Patay, G., Egri, G., Häggkvist, J., Varrone, A., Farde, L., Halldin, C., and Gulyás, B. (2013). Performance Evaluation of the Small-Animal nanoScan PET/MRI System. *Journal of Nuclear Medicine*, 54(10):1825–1832. 87

Nakamoto, Y., Osman, M., Cohade, C., Marshall, L. T., Links, J. M., Kohlmyer, S., and Wahl, R. L. (2002). PET/CT: comparison of quantitative tracer uptake between germanium and CT transmission attenuation-corrected images. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 43(9):1137–43. 32, 67

National Electrical Manufacturers Association (2008). *NEMA Standards Publication NU 4-2008 Performance Measurements of Small Animal Positron Emission Tomographs*. National Electrical Manufacturers Association, Rosslyn, VA. xxii, xxiii, xxvi, xxvii, 21, 55, 56, 57, 98, 128, 130, 131, 154, 171, 176

Niu, X., Dong, Y., Ye, H., and Wang, W. (2012). A study of count-rate dependent normalization effect in 3D-PET data. In *IEEE Nuclear Science Symposium Conference Record*, pages 2991–2994. IEEE. 30

Nutt, R. (2002). The History of Positron Emission Tomography. *Molecular Imaging & Biology*, 4(1):11–26. xvii, 1

Nuyts, J., Beque, D., Dupont, P., and Mortelmans, L. (2002). A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Transactions on Nuclear Science*, 49(1):56–60. xviii, xix, xxi, xxvi, 23, 28, 62, 68, 69, 84

NVIDIA, editor (2021). *CUDA Fortran Programming Guide*. NVIDIA. 45

Oliver, J. F. and Rafecas, M. (2008). Revisiting the singles rate method for modeling accidental coincidences in PET. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 4288–4291. IEEE. 34

Ollinger, J. M. (1996). Model-based scatter correction for fully 3D PET. *Physics in Medicine and Biology*, 41(1):153–176. xxiii, xxvii, 33, 136

Onecha, V. V., Galve, P., Ibanez, P., Espinosa, A., Sanchez-Parcerisa, D., Espana, S., Fraile, L. M., and Udias, J. M. (2020). Dictionary based MLEM-algorithm for real-time proton range verification from PET data: the virtue of contrasts. In *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. 108

Panin, V., Kehren, F., Michel, C., and Casey, M. (2006). Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE Transactions on Medical Imaging*, 25(7):907–921. 25, 161

Papathanassiou, D., Bruna-Muraille, C., Liehn, J.-C., Nguyen, T. D., and Curé, H. (2009). Positron Emission Tomography in oncology: Present and future of PET and PET/CT. *Critical Reviews in Oncology/Hematology*, 72(3):239–254. 1

Paredes-Pacheco, J., López-González, F. J., Silva-Rodríguez, J., Efthimiou, N., Niñerola-Baizán, A., Ruibal, Á., Roé-Vellvé, N., and Aguiar, P. (2021). SimPET - An open online platform for the Monte Carlo simulation of realistic brain PET data. Validation for 18 F-FDG scans. *Medical Physics*, page mp.14838. 108

Park, S.-J., Rogers, W. L., and Clinthorne, N. H. (2007). Effects of Positron Range and Annihilation Photon Acolinearity on Image Resolution of a Compton PET. *IEEE Transactions on Nuclear Science*, 54(5):1543–1552. 4, 118

Peng, H. and Levin, C. S. (2012). Study of PET intrinsic spatial resolution and contrast recovery improvement for PET/MRI systems. *Physics in Medicine and Biology*, 57(9):N101–N115. 4, 118

Peng, P., Judenhofer, M. S., Jones, A. Q., and Cherry, S. R. (2018). Compton PET: a simulation study for a PET module with novel geometry and machine learning for position decoding. *Biomedical Physics & Engineering Express*, 5(1):015018. 107

Perez-Benito, D., Chil, R., Udias, J., Desco, M., and Vaquero, J. (2018). SiPM-based PET detector module for a 4π span scanner. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. xvii, 10, 61, 87, 112, 144

Pfaehler, E., De Jong, J. R., Dierckx, R. A. J. O., van Velden, F. H. P., and Boellaard, R. (2018). SMART (SiMulAtion and ReconsTruction) PET: an efficient PET simulation-reconstruction tool. *EJNMMI Physics*, 5(1):16. 109

Phelps, M., Hoffman, E., Huang, S., and Kuhl, D. (1977). ECAT: a new computerized tomographic imaging system for position-emitting radiopharmaceuticals. Technical Report 6, Oakland Operations Office, Oakland, CA. xvii, xxi, xxv

Pilleri, A., Camarlinghi, N., Del Guerra, A., Sportelli, G., and Belcari, N. (2019). A Monte Carlo detector response model for the IRIS PET preclinical scanner. *Physica Medica*, 57(December 2018):107–114. 25, 61, 148, 161

Polycarpou, I., Thielemans, K., Manjeshwar, R., Aguiar, P., Marsden, P. K., and Tsoumpas, C. (2011). Comparative evaluation of scatter correction in 3D PET using different scatter-level approximations. *Annals of Nuclear Medicine*, 25(9):643–649. 136

Pönisch, F., Enghardt, W., and Lauckner, K. (2003). Attenuation and scatter correction for in-beam positron emission tomography monitoring of tumour irradiations with heavy ions. *Physics in Medicine and Biology*, 48(15):2419–2436. 33, 136

Poon, J. K., Dahlbom, M. L., Casey, M. E., Qi, J., Cherry, S. R., and Badawi, R. D. (2015). Validation of the SimSET simulation package for modeling the Siemens Biograph mCT PET scanner. *Physics in Medicine and Biology*, 60(3):N35–N45. 108

Popota, F. D., Aguiar, P., España, S., Lois, C., Udias, J. M., Ros, D., Pavia, J., and Gispert, J. D. (2015). Monte Carlo simulations versus experimental measurements in a small animal PET system. A comparison in the NEMA NU 4-2008 framework. *Physics in Medicine and Biology*, 60(1):151–162. 49

Pratx, G., Chinn, G., Olcott, P., and Levin, C. (2009). Fast, Accurate and Shift-Varying Line Projections for Iterative Reconstruction Using the GPU. *IEEE Transactions on Medical Imaging*, 28(3):435–445. 24, 44, 61, 83

Pratx, G. and Xing, L. (2011). GPU computing in medical physics: A review. *Medical Physics*, 38(5):2685–2697. 44

Qi, J. and Leahy, R. M. (2006). Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine and Biology*, 51(15):R541–R578. xviii, 18, 23, 26, 27, 28

Qi, J., Yang, Y., Zhou, J., Wu, Y., and Cherry, S. R. (2011). Experimental assessment of resolution improvement of a zoom-in PET. *Physics in Medicine and Biology*, 56(17):N165–N174. xvii, 10, 32, 87, 144

Qian, H., Rui, X., and Ahn, S. (2017). Deep Learning Models for PET Scatter Estimations. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–5. IEEE. 33, 36, 137

Rafecas, M., Mosler, B., Dietz, M., Pogl, M., Stamatakis, A., McElroy, D., and Ziegler, S. (2004). Use of a Monte Carlo-based probability matrix for 3-D iterative reconstruction of

MADPET-II data. *IEEE Transactions on Nuclear Science*, 51(5):2597–2605. 25, 107, 148, 160, 161

Rapisarda, E., Bettinardi, V., Thielemans, K., and Gilardi, M. C. (2010). Image-based point spread function implementation in a fully 3D OSEM reconstruction algorithm for PET. *Physics in Medicine and Biology*, 55(14):4131–4151. 25, 83, 147

Reader, A., Julyan, P., Williams, H., Hastings, D., and Zweit, J. (2003). EM algorithm system modeling by image-space techniques for PET reconstruction. *IEEE Transactions on Nuclear Science*, 50(5):1392–1397. 24, 62, 63, 64

Reader, A. J., Corda, G., Mehranian, A., da Costa-Luis, C., Ellis, S., and Schnabel, J. A. (2021). Deep Learning for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):1–25. xviii, 18, 27, 28, 29, 36

Reader, A. J., Letourneau, E., and Verhaeghe, J. (2011). Generalization of the image space reconstruction algorithm. In *2011 IEEE Nuclear Science Symposium Conference Record*, pages 4233–4238. IEEE. xviii, 22

Reader, A. J. and Schramm, G. (2021). Artificial Intelligence for PET Image Reconstruction. *Journal of Nuclear Medicine*, 62(10):1330–1333. xviii, 18, 36

Rezaei, A., Defrise, M., Bal, G., Michel, C., Conti, M., Watson, C., and Nuyts, J. (2012). Simultaneous Reconstruction of Activity and Attenuation in Time-of-Flight PET. *IEEE Transactions on Medical Imaging*, 31(12):2224–2233. 32

Rezaei, A., Defrise, M., and Nuyts, J. (2014). ML-Reconstruction for TOF-PET With Simultaneous Estimation of the Attenuation Factors. *IEEE Transactions on Medical Imaging*, 33(7):1563–1572. 32

Rodriguez, M., Liow, J.-S., Thada, S., Sibomana, M., Chelikani, S., Mulnix, T., Johnson, C. A., Michel, C., Barker, W. C., and Carson, R. E. (2007). Count-Rate Dependent Component-Based Normalization for the HRRT. *IEEE Transactions on Nuclear Science*, 54(3):486–495. 30, 35

Rogasch, J. M., Boellaard, R., Pike, L., Borchmann, P., Johnson, P., Wolf, J., Barrington, S. F., and Kobe, C. (2021). Moving the goalposts while scoring—the dilemma posed by new PET technologies. *European Journal of Nuclear Medicine and Molecular Imaging*. 24, 84

Rogasch, J. M., Hofheinz, F., Lougovski, A., Furth, C., Ruf, J., Großer, O. S., Mohnike, K., Hass, P., Walke, M., Amthauer, H., and Steffen, I. G. (2014). The influence of different signal-to-background ratios on spatial resolution and F18-FDG-PET quantification using point spread function and time-of-flight reconstruction. *EJNMMI Physics*, 1(1):12. 84

Rogers, D. (1984). Low energy electron transport with EGS. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 227(3):535–548. 108

Rogers, D. W. O. (2006). Fifty years of Monte Carlo simulations for medical physics. *Physics in Medicine and Biology*, 51(13):R287–R301. 109

Salvat, F. (1987). Algorithms for random sampling from single-variate distributions. *Computer Physics Communications*, 46(3):427–436. 42

Samanta, S., Jiang, J., Hamdi, M., Register, A. Z., Majewski, S., Williams, M. B., Turkington, T. G., Tornai, M. P., Laforest, R., O'Sullivan, J. A., and Tai, Y.-C. (2021). Performance comparison of a dedicated total breast PET system with a clinical whole-body PET system: a simulation study. *Physics in Medicine and Biology*, 66(11):115004. xvii, 10, 32, 61, 144

Sanders, J. and Kandrot, E. (2010). *The CUDA Handbook: A Comprehensive Guide to GPU Programming.* Addison-Wesley Professional, 1st edition. 45

Sari, H., Reaungamornrat, J., Catalano, O., Vera-Olmos, J., Izquierdo-Garcia, D., Morales, M. A., Torrado-Carvajal, A., Ng, S. C. T., Malpica, N., Kamen, A., and Catana, C. (2021). Evaluation of Deep Learning-based Approaches to Segment Bowel Air Pockets and Generate Pelvis Attenuation Maps from CAIPIRINHA-accelerated Dixon MR Images. *Journal of Nuclear Medicine*, page jnumed.120.261032. 36

Sarrut, D., Bała, M., Bardiès, M., Bert, J., Chauvin, M., Chatzipapas, K., Dupont, M., Etxebeste, A., Fanchon, L. M., Jan, S., Kayal, G., Kirov, A. S., Kowalski, P., Krzemien, W., Labour, J., Lenz, M., Loudos, G., Mehadji, B., Ménard, L., Morel, C., Papadimitroulas, P., Rafecas, M., Salvadori, J., Seiter, D., Stockhoff, M., Testa, E., Trigila, C., Pietrzyk, U., Vandenberghe, S., Verdier, M.-A., Visvikis, D., Ziemons, K., Zvolský, M., and Roncali, E. (2021). Advanced Monte Carlo simulations of emission tomography imaging systems with GATE. *Physics in Medicine and Biology*, 66(10):10TR03. 108

Sarrut, D., Bardiès, M., Boussion, N., Freud, N., Jan, S., Létang, J.-M., Loudos, G., Maigne, L., Marcatili, S., Mauxion, T., Papadimitroulas, P., Perrot, Y., Pietrzyk, U., Robert, C., Schaart, D. R., Visvikis, D., and Buvat, I. (2014). A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications. *Medical Physics*, 41(6Part1):064301. 108

Scheins, J. J., Herzog, H., and Shah, N. J. (2011). Fully-3D PET Image Reconstruction Using Scanner-Independent, Adaptive Projection Data and Highly Rotation-Symmetric Voxel Assemblies. *IEEE Transactions on Medical Imaging*, 30(3):879–892. 61

Schneider, W., Bortfeld, T., and Schlegel, W. (2000). Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. *Physics in Medicine and Biology*, 45(2):459–478. 32, 67, 124

Schramm, G., Rigie, D., Vahle, T., Rezaei, A., Van Laere, K., Shepherd, T., Nuyts, J., and Boada, F. (2021). Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network. *NeuroImage*, 224(February 2020):117399. 36

Sempau, J., Wilderman, S. J., and Bielajew, A. F. (2000). DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations. *Physics in Medicine and Biology*, 45(8):2263–2291. 109

Shepp, L. A. and Vardi, Y. (1982). Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122. xviii, 22, 25, 26, 67, 105, 150

Shiri, I., Arabi, H., Geramifar, P., Hajianfar, G., Ghafarian, P., Rahmim, A., Ay, M. R., and Zaidi, H. (2020). Deep-JASC: joint attenuation and scatter correction in whole-body 18F-FDG PET using a deep residual network. *European Journal of Nuclear Medicine and Molecular Imaging*, 47(11):2533–2548. 33, 36, 137

Siddon, R. L. (1985). Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12(2):252–255. 24, 84

Sidky, E. Y. and Pan, X. (2008). Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology*, 53(17):4777–4807. 23

Simonyan, K. and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14. 35, 36

Skansi, S. (2018). *Introduction to Deep Learning.* Undergraduate Topics in Computer Science. Springer International Publishing, Cham. 35, 36

Slart, R. H. J. A., Tsoumpas, C., Glaudemans, A. W. J. M., Noordzij, W., Willemsen, A. T. M., Borra, R. J. H., Dierckx, R. A. J. O., and Lammertsma, A. A. (2021). Long axial field of view PET scanners: a road map to implementation and new possibilities. *European Journal of Nuclear Medicine and Molecular Imaging*, 48(13):4236–4245. xvii, 61

Slomka, P. J., Pan, T., and Germano, G. (2016). Recent Advances and Future Progress in PET Instrumentation. *Seminars in Nuclear Medicine*, 46(1):5–19. xvii, 13

Southekal, S., Purschke, M. L., Schlyer, D. J., and Vaska, P. (2011). Quantitative PET Imaging Using a Comprehensive Monte Carlo System Model. *IEEE Transactions on Nuclear Science*, 58(5):2286–2295. 25, 107, 148, 161

Stearns, C., McDaniel, D., Kohlmyer, S., Arul, P., Geiser, B., and Shanmugam, V. (2003). Random coincidence estimation from single event rates on the Discovery ST PET/CT scanner.

In *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No.03CH37515)*, volume 5, pages 3067–3069. IEEE. 34

Stockhoff, M., Decuyper, M., Van Holen, R., and Vandenberghe, S. (2021). High-resolution monolithic LYSO detector with 6-layer depth-of-interaction for clinical PET. *Physics in Medicine & Biology*, 66(15):155014. xvii, 12

Surti, S. and Karp, J. S. (2016). Advances in time-of-flight PET. *Physica Medica*, 32(1):12–22. xvii, 11

Tao, W., Chen, G., Weng, F., Zan, Y., Zhao, Z., Peng, Q., Xu, J., and Huang, Q. (2018). Simulation study of a high-performance brain PET system with dodecahedral geometry. *Medical Physics*, 45(7):3297–3304. xvii, 61, 144

Tao, W., Weng, F., Chen, G., Lv, L., Zhao, Z., Xie, S., Zan, Y., Xu, J., Huang, Q., and Peng, Q. (2020). Design study of fully wearable high-performance brain PETs for neuroimaging in free movement. *Physics in Medicine and Biology*, 65(13). 107

Theodorakis, L., Loudos, G., Prassopoulos, V., Kappas, C., Tsougos, I., and Georgoulias, P. (2013). A review of PET normalization. *Nuclear Medicine Communications*, 34(11):1033–1045. 30, 66

Thielemans, K., Manjeshwar, R., Tsoumpas, C., and Jansen, F. (2007). A new algorithm for scaling of PET scatter estimates using all coincidence events. In *2007 IEEE Nuclear Science Symposium Conference Record*, volume 5, pages 3586–3590. IEEE. 33, 136

Thielemans, K., Tsoumpas, C., Mustafovic, S., Beisel, T., Aguiar, P., Dikaios, N., and Jacobson, M. W. (2012). STIR: software for tomographic image reconstruction release 2. *Physics in Medicine and Biology*, 57(4):867–883. 109

Thompson, C. J., Moreno-Cantu, J., and Picard, Y. (1992). PETSIM: Monte Carlo simulation of all sensitivity and resolution parameters of cylindrical positron imaging systems. *Physics in Medicine and Biology*, 37(3):731–749. 108

Tohme, M. S. and Qi, J. (2009). Iterative image reconstruction for positron emission tomography based on a detector response function estimated from point source measurements. *Physics in Medicine and Biology*, 54(12):3709–3725. 25, 88, 147, 161

Tong, S., Alessio, A. M., and Kinahan, P. E. (2010). Image reconstruction for PET/CT scanners: past achievements and future challenges. *Imaging in Medicine*, 2(5):529–545. xviii, 18, 19, 23, 27

Torrado-Carvajal, A., Vera-Olmos, J., Izquierdo-Garcia, D., Catalano, O. A., Morales, M. A., Margolin, J., Soricelli, A., Salvatore, M., Malpica, N., and Catana, C. (2019). Dixon-VIBE Deep Learning (DIVIDE) Pseudo-CT Synthesis for Pelvis PET/MR Attenuation Correction. *Journal of Nuclear Medicine*, 60(3):429–435. 32, 36

Tuna, U., Peltonen, S., and Ruotsalainen, U. (2009). Interpolation for the gap-filling of the HRRT PET sinograms by using the slices in the direction of the radial samples. In *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 3273–3279. IEEE. 88

Udias, J. M., Fernandez, C. G., Herraiz, J. L., Pérez-Benito, D., Galve, P., Lopez-Montes, A., Lopez-Longas, J., Arco, J. M., Desco, M., and Vaquero, J. J. (2018). Performance evaluation of the PET subsystem of the extended FOV SuperArgus 6R preclinical scanner. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. xvii, xix, xxii, xxiii, xxvi, xxvii, 27, 32, 61, 82, 113, 115, 127

Ulam, S., Richtmeyer, R. D., and von Neumann, J. (1947). Statistical Methods in Neutron Diffusion. 108

Valladolid Onecha, V., Galve, P., Arias-Valcayo, F., Freijo Escudero, C., Sanchez-Parcerisa, D., Ibanez, P., Espana, S., Fraile, L. M., and Udias, J. M. (2021). Dictionary based MLEM & Simulated Annealing (MSA) GPU-Algorithm for real time PET proton range verification. In *2021 Virtual IEEE Nuclear Science Symposium and Medical Imaging Conference*. 108

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, 2(1):e453. 93

van Gómez López, O. (2021). *Tumor heterogeneity in PET-CT images*. PhD thesis, Universidad Complutense de Madrid. xviii

Vandenberghe, S., Mikhaylova, E., D'Hoe, E., Mollet, P., and Karp, J. S. (2016). Recent developments in time-of-flight PET. *EJNMMI Physics*, 3(1):3. 11

Vaquero, J., Sánchez, J., Udías, J., Cal-González, J., and Desco, M. (2013). MRI compatibility of position-sensitive photomultiplier depth-of-interaction PET detectors modules for in-line multimodality preclinical studies. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 702:83–87. 13

Vicente, E. (2012). *Characterization, improvement and design of preclinical PET scanners*. PhD thesis, Universidad Complutense de Madrid. xviii, 35

Vicente, E., Herraiz, J. L., España, S., Herranz, E., Desco, M., Vaquero, J. J., and Udías, J. M. (2013). Improved dead-time correction for PET scanners: application to small-animal PET. *Physics in Medicine and Biology*, 58(7):2059–2072. xviii, 34

Vicente, E., Soto-Montenegro, M., Espana, S., Herraiz, J. L., Herranz, E., Vaquero, J. J., Desco, M., and Udias, J. M. (2007). Influence of random, pile-up and scatter corrections in the quantification properties of small-animal PET scanners. In *2007 IEEE Nuclear Science Symposium Conference Record*, volume 5, pages 3964–3968. IEEE. xviii, 35

Vicente, E., Vaquero, J. J., Espana, S., Herraiz, J. L., Udias, J. M., and Desco, M. (2006). Normalization in 3D PET: Dependence on the Activity Distribution of the Source. In *2006 IEEE Nuclear Science Symposium Conference Record*, volume 4, pages 2206–2209. IEEE. 30

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, , Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272. 93

Wagenknecht, G., Kaiser, H.-J., Mottaghy, F. M., and Herzog, H. (2013). MRI for attenuation correction in PET: methods and challenges. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 26(1):99–113. 32

Walker, A. J. (1977). An Efficient Method for Generating Discrete Random Variables with General Distributions. *ACM Transactions on Mathematical Software*, 3(3):253–256. 42

Wang, G. and Qi, J. (2015). PET Image Reconstruction Using Kernel Method. *IEEE Transactions on Medical Imaging*, 34(1):61–71. 27

Wang, Y., Seidel, J., Tsui, B. M. W., Vaquero, J. J., and Pomper, M. G. (2006). Performance evaluation of the GE healthcare eXplore VISTA dual-ring small-animal PET scanner. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 47(11):1891–900. 10, 17, 27, 32, 51, 87, 112, 115, 144

Watson, C. (2000). New, faster, image-based scatter correction for 3D PET. *IEEE Transactions on Nuclear Science*, 47(4):1587–1594. xxiii, xxvii, 33, 136

Watson, C. C., Newport, D., and Casey, M. E. (1996). A Single Scatter Simulation Technique for Scatter Correction in 3D PET. In *Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 255–268. xxiii, xxvii, 33, 136

Wei, S. and Vaska, P. (2020). Evaluation of quantitative, efficient image reconstruction for VersaPET, a compact PET system. *Medical Physics*, 47(7):2852–2868. 25, 61, 107, 148, 161

Weishi Xia, Lewitt, R., and Edholm, P. (1995). Fourier correction for spatially variant collimator blurring in SPECT. *IEEE Transactions on Medical Imaging*, 14(1):100–115. 18

Werling, A., Bublitz, O., Doll, J., Adam, L.-E., and Brix, G. (2002). Fast implementation of the single scatter simulation algorithm and its use in iterative image reconstruction of PET data. *Physics in Medicine and Biology*, 47(16):310. 33, 136

Werner, C. J., Armstrong, J., Brown, F. B., Bull, J. S., Casswell, L., Cox, L. J., Dixon, D., Forster, R. A., Goorley, J. T., Hughes, H. G., Favorite, J., Martz, R., Mashnik, S. G., Rising, M. E., Solomon, C., Sood, A., Sweezy, J. E., Werner, C. J., Zukaitis, A., Anderson, C., Elson, J. S., Durkee, J. W., Johns, R. C., McKinney, G. W., McMath, G. E., Hendricks, J. S., Pelowitz, D. B., Prael, R. E., Booth, T. E., James, M. R., Fensin, M. L., Wilcox, T. A., and Kiedrowski, B. C. (2017). MCNP User's Manual Code Version 6.2. *Los Alamos National Laboratory.* 108

Whiteley, W., Luk, W. K., and Gregor, J. (2020). DirectPET: full-size neural network PET reconstruction from sinogram data. *Journal of Medical Imaging*, 7(03):1. 36

Woodcock, E., Murphy, T., Hemmings, P., and Longworth, S. (1965). Techniques used in the GEM code for Monte Carlo neutronics calculation. *Proc. Conf. Applications of Computing Methods to Reactors, ANL-7050*, (3):557–579. 123

Xiang, H., Lim, H., Fessler, J. A., and Dewaraja, Y. K. (2020). A deep neural network for fast and accurate scatter estimation in quantitative SPECT/CT under challenging scatter conditions. *European Journal of Nuclear Medicine and Molecular Imaging*, 47(13):2956–2967. 33, 36, 137

Xu, H., Lenz, M., Caldeira, L., Bo, M., Pietrzyk, U., Lerche, C. W., Shah, N. J., and Scheins, J. J. (2019). Resolution modeling in projection space using a factorized multi-block detector response function for PET image reconstruction. *Physics in Medicine and Biology.* 25, 61, 148, 161

Xuan Liu, Defrise, M., Michel, C., Sibomana, M., Comtat, C., Kinahan, P., and Townsend, D. (1999). Exact rebinning methods for three-dimensional PET. *IEEE Transactions on Medical Imaging*, 18(8):657–664. 18, 20

Yang, J., Park, D., Gullberg, G. T., and Seo, Y. (2019). Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain 18 F-FDG PET. *Physics in Medicine & Biology*, 64(7):075019. 33, 36, 87, 137

Yang, J., Sohn, J. H., Behr, S. C., Gullberg, G. T., and Seo, Y. (2021). CT-less Direct Correction of Attenuation and Scatter in the Image Space Using Deep Learning for Whole-Body FDG

PET: Potential Benefits and Pitfalls. *Radiology: Artificial Intelligence*, 3(2):e200137. 33, 36, 137

Ye, J., Song, X., and Hu, Z. (2014). Scatter correction with combined single-scatter simulation and Monte Carlo simulation for 3D PET. In *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3. IEEE. 33, 107, 136, 137, 146

Ye, T., Chai, P., Gao, J., Yun, M.-K., Liu, S.-Q., Shan, B.-C., and Wei, L. (2011). Investigation of scatter from out of the field of view and multiple scatter in PET using Monte Carlo simulations. *Chinese Physics C*, 35(12):1166–1171. 32, 136

Yin, T. and Obi, T. (2021). Generation of attenuation correction factors from time-of-flight PET emission data using high-resolution residual U-net. *Biomedical Physics & Engineering Express*, 7(6):065006. 36

Yoshida, E., Tashima, H., Akamatsu, G., Iwao, Y., Takahashi, M., Yamashita, T., and Yamaya, T. (2020). 245 ps-TOF brain-dedicated PET prototype with a hemispherical detector arrangement. *Physics in Medicine and Biology*, 65(14):145008. xvii, 10, 61, 144

Zaidi, H. and Karakatsanis, N. (2018). Towards enhanced PET quantification in clinical oncology. *The British Journal of Radiology*, 91(1081):20170508. 18, 24, 27, 29, 147

Zaidi, H., Labbé, C., and Morel, C. (1998). Implementation of an environment for Monte Carlo simulation of fully 3-D positron tomography on a high-performance parallel platform. *Parallel Computing*, 24(9-10):1523–1536. 108

Zeng, G., Gullberg, G., Tsui, B., and Terry, J. (1991). Three-dimensional iterative reconstruction algorithms with attenuation and geometric point response correction. *IEEE Transactions on Nuclear Science*, 38(2):693–702. 24, 88, 104

Zhang, L., Staelens, S., Van Holen, R., De Beenhouwer, J., Verhaeghe, J., Kawrakow, I., and Vandenberghe, S. (2010). Fast and memory-efficient Monte Carlo-based image reconstruction for whole-body PET. *Medical Physics*, 37(7):3667–3676. 25, 107, 148, 160, 161

Zhou, J. and Qi, J. (2011). Fast and efficient fully 3D PET image reconstruction using sparse system matrix factorization with GPU acceleration. *Physics in Medicine and Biology*, 56(20):6739–6757. 25, 61, 106, 147, 161

Zhou, J. and Qi, J. (2014). Efficient fully 3D list-mode TOF PET image reconstruction using a factorized system matrix with an image domain resolution model. *Physics in Medicine and Biology*, 59(3):541–559. 25, 61, 83, 148, 161

Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R., and Rosen, M. S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492. 36

Zhu, X. and Fakhri, G. E. (2013). Proton Therapy Verification with PET Imaging. *Theranostics*, 3(10):731–740. 107

Scientific publications and conference contributions

During this PhD. Thesis the following investigations were published in international journals relevant in the field or presented in international conferences.

Scientific publications

1. **Galve P**, Udias J M, Lopez-Montes A, Arias-Valcayo F, Vaquero J J, Desco M and Herraiz J L 2021 *Super-Iterative Image Reconstruction in PET* ed S Matej and S D Metzler IEEE Trans. Comput. Imaging 7 248–57
   Online: https://doi.org/10.1109/TCI.2021.3059107
2. Piersa-Siłkowska M, et al. 2021 *First β-decay spectroscopy of ^{135}In and new β-decay branches of ^{134}In* Phys. Rev. C 104 044328
   Online: https://doi.org/10.1103/PhysRevC.104.044328
3. López-Montes A, **Galve P**, Udias J M, Cal-González J, Vaquero J J, Desco M and Herraiz J L 2020 *Real-Time 3D PET Image with Pseudoinverse Reconstruction* Appl. Sci. 10 2829
   Online: https://doi.org/10.3390/app10082829
4. Benito J, et al. 2020 *Detailed spectroscopy of doubly magic ^{132}Sn* Phys. Rev. C 102 014328
   Online: https://doi.org/10.1103/PhysRevC.102.014328

Conference contributions and proceedings

1. **Galve P**, Arias-Valcayo F, Lopez-Montes A, Villa-Abaunza A, Ibanez P, Herraiz J L and Udias J M 2021 *Ultra-fast Fully Monte Carlo PET Reconstruction* 2021 Virtual IEEE Nuclear Science Symposium and Medical Imaging Conference
   Online: https://nssmic.ieee.org/2021/

2. Valladolid Onecha V, **Galve P**, Arias-Valcayo F, Freijo Escudero C, Sanchez-Parcerisa D, Ibanez P, Espana S, Fraile L M and Udias J M 2021 *Dictionary based MLEM & Simulated Annealing (MSA) GPU-Algorithm for real time PET proton range verification* 2021 Virtual IEEE Nuclear Science Symposium and Medical Imaging Conference
   Online: https://nssmic.ieee.org/2021/
3. **Galve P**, Arias-Valcayo F, Lopez-Montes A, Villa-Abaunza A, Ibanez P, Herraiz J L and Udias J M 2021 *Ultra-fast Monte Carlo PET Reconstructor* 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine pp 152–6.
   Online: https://arxiv.org/abs/2110.04143
4. **Galve P**, Herraiz J L, Catana C and Udias J M 2020 *GPU based Fast and Flexible Iterative Reconstructions of Arbitrary and Complex PET Scanners: Application to Next Generation Dedicated Brain Scanners* 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)
   Online: https://nssmic.ieee.org/2020/
5. **Galve P**, Arias-Valcayo F, Lopez-Montes A, Villa-Abaunza A, Ibanez P, Herraiz J L and Udias J M 2020 *Multi-purpose Ultra-fast Monte Carlo PET simulator* 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)
   Online: https://nssmic.ieee.org/2020/
6. Onecha V V, **Galve P**, Ibanez P, Espinosa A, Sanchez-Parcerisa D, Espana S, Fraile L M and Udias J M 2020 *Dictionary based MLEM-algorithm for real-time proton range verification from PET data: the virtue of contrasts* 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)
   Online: https://nssmic.ieee.org/2020/
7. Arias-Valcayo F, Herraiz J L, Lopez-Montes A, Vaquero J J, Desco M, **Galve P** and Udias J M 2020 *No gates , Real-Time Cardiac Reconstruction of Small Animal PET Data* 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)
   Online: https://nssmic.ieee.org/2020/
8. Lopez-Montes A, Herraiz J L, **Galve P**, Espana S, Vicente E, Cal-Gonzalez J and Udias J M 2019 *PeneloPET v3.0, an improved multiplatform PET Simulator* 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (IEEE) pp 1–3
   Online: https://doi.org/10.1109/NSS/MIC42101.2019.9059837
9. **Galve P**, López-Montes A, Udías J M and López Herraiz J 2019 *Iterative Algorithm for Optimal Super Resolution Sampling* Springer Proceedings in Physics vol 225 pp 141–3
   Online: https://doi.org/10.1007/978-3-030-22204-8_10
10. López-Montes A, **Galve P**, Udías J M and López Herraiz J 2019 *Real-Time Tomographic Image Reconstruction in PET Using the Pseudoinverse of the System Response Matrix* Springer Proceedings in Physics vol 225 pp 207–9
    Online: https://doi.org/10.1007/978-3-030-22204-8_10

11. **Galve P**, Udias J M, Lopez-Montes A and Herraiz J L 2019 *Super-iterative image reconstruction in PET* 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine ed S Matej and S D Metzler (SPIE) p 53
    Online: https://doi.org/10.1117/12.2534760
12. Lopez Montes A, **Galve P**, Udias J M and Herraiz J L 2019 *Application of the pseudoinverse for real-time 3D PET image reconstruction* 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine ed S Matej and S D Metzler (SPIE) p 74
    Online: https://doi.org/10.1117/12.2534841
13. Arias-Valcayo F, Herraiz J L, **Galve P**, Vaquero J J, Desco M and Udias J M 2019 *Awake preclinical brain PET imaging based on point sources* 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine ed S Matej and S D Metzler (SPIE) p 67
    Online: https://doi.org/10.1117/12.2534830
14. **Galve P**, Herraiz J L, Lopez-Montes A, Gutierrez C and Udias J M 2018 *Effect of Incomplete Acquisitions on the Image of High Resolution Preclinical Scanners* 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC) (Sidney, Australia)
    Online: http://www.nssmic.org/2018/
15. Udias J M, Fernandez C G, Herraiz J L, Pérez-Benito D, **Galve P**, Lopez-Montes A, Lopez-Longas J, Arco J M, Desco M and Vaquero J J 2018 *Performance evaluation of the PET subsystem of the extended FOV SuperArgus 6R preclinical scanner* 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)
    Online: http://www.nssmic.org/2018/
16. **Galve P**, Lopez-Montes A, Gutierrez Fernandez C, Herraiz J L and Udias J M 2018 *Effect of incomplete acquisitions on the image of high resolution preclinical scanners* II Jornadas RSEF/IFIMED de Física Médica
    Online: https://indico.ific.uv.es/event/3295/overview
17. **Galve P**, Lopez-Montes A, Udias J M, Moore S C and Herraiz J L 2017 *Resolution Recovery in PET using Maximum Likelihood-based Sub-Crystals* XXXVI Reunión Bienal y 27º Encuentro Ibérico de Enseñanza de la Física ed D Cotina Gil and E López Lago (Santiago de Compostela: Real Sociedad Española de Física) pp 354–5
    Online: https://indico.ific.uv.es/event/2851/overview
18. Lopez-Montes A, **Galve P**, Udias J M and Herraiz J L 2017 *Real-Time PET Data Axial Rebinning Based on the Pseudoinversion of the System Response Matrix* XXXVI Reunión Bienal y 27º Encuentro Ibérico de Enseñanza de la Física ed D Cotina Gil and E López Lago (Santiago de Compostela: Real Sociedad Española de Física) pp 358–9
    Online: https://indico.ific.uv.es/event/2851/overview

19. Herraiz J L, **Galve P**, Mahmood U and Udias J M 2017 *Simulation of Simultaneous PET Imaging of Antibodies Labeled with Zr-89 and I-124 Based on Triple Coincidences* XXXVI Reunión Bienal y 27º Encuentro Ibérico de Enseñanza de la Física ed D Cotina Gil and E López Lago (Santiago de Compostela: Real Sociedad Española de Física) pp 361–2
    Online: https://indico.ific.uv.es/event/2851/overview
20. **Galve P**, Lopez-Montes A, Udias J M, Moore S C and Herraiz J L 2017 *Data-driven Improved Sampling in PET* 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (IEEE) pp 1–5
    Online: https://doi.org/10.1109/NSSMIC.2017.8532745
21. Lopez-Montes A, **Galve P**, Udias J M and Herraiz J L 2017 *Real-Time Accurate Rebinning of PET Data Based on the Pseudo-Inverse of the Axial System Matrix* 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) vol 2 (IEEE) pp 1–4
    Online: https://doi.org/10.1109/NSSMIC.2017.8532880
22. **Galve P**, Herraiz J L and Udias J M 2016 *Estrategias de corrección por profundidad de interacción en escáneres PET preclínico de alta sensibilidad* VIII Jornadas CPAN
    Online: https://www.i-cpan.es/jornadas8/https://www.i-cpan.es/jornadas8/